

25 February 1982

What to tell the White House

The new White House Science Council is only a pale shadow of the old President's Science Advisory Committee; it should stay that way.

The King is dead; long live the King (or the Queen as the case may be)! Can this declaration, much cherished by historians keen to demonstrate the repetitiveness of events, be applied to that curious political animal in the United States — the President's Science Advisory Committee? The committee (PSAC, pronounced *pea-sack*) was abolished in 1973 by President Nixon, as legend has it because it had become too uppish. Although many in the then administration, including even the Secretary of State, Dr Henry Kissinger, agreed with PSAC's most talkative members that anti-ballistic missile systems were then technically backward and always likely to be strategically ineffectual (and, for that reason, dangerous), the administration plainly did not want its negotiating position compromised by a bunch of fast and independently-talking academics, among whom members of PSAC were prominent. But now, after a decent interval, there is to be a science advisory council in the White House again (see page 637). PSAC having ceased to exist, it seems as if it has been necessary to reinvent it as the White House Science Council. Appearances, however, are misleading.

In its heyday, perhaps until the mid-1960s, PSAC was a remarkable institution. In many ways, it was a continuation into peacetime (made to seem threatening by the arrival of the first sputnik) of the camaraderie of the veterans of technological wars — the people who had set up radiation laboratories, Manhattan Projects and the like. PSAC in its early days was immensely influential because its members were influential, both in the academic community and with the government. PSAC seemed merely to have to wave its wand, and the federal government would agree that there should be a programme of curriculum development aimed at the improvement of science teaching in high schools, for example — ironically, the programme now finally dismantled in the Administration's budget for 1983. On other occasions, the committee lent powerful support, from outside the White House, to the executive functions of the Science Advisor as, for example, in the early 1960s when Dr Jerome Wiesner was battling with Mr McGeorge Bundy for an influence over United States policy on disarmament (and defence). Unlike most other advisory committees, PSAC seems not to have bothered to wait for a request for advice but would offer its opinions gratuitously, sometimes to people who did not want to hear. Always, PSAC gave the technical community in the United States the sense, but also the illusion, that there was an effective and publicly constituted channel by means of which account could be taken of raw technical opinion. What went wrong?

The doctrine that everything that President Nixon did must by definition be mistaken is a poor guide to the truth. While the Nixon administration may have been more afflicted by public disagreement about arms control policy than previous administrations, there is also something in the view that the virtues that made PSAC seem to the technical community to be exceedingly important were precisely those calculated to be a thorn in the flesh of the White House. The dilemma has been neatly described by Dr Edward J. Burger in his book *Science at the White House*, published last year. (For Professor Harvey Brooks's review, see *Nature*, 290, 635; 1981). Because many of the committee's pronouncements were unsolicited, they were not assured of the welcome they might otherwise have had. But when asked for its advice, PSAC would often provide its masters with

an opinion which, even if entirely correct, could only seem to the embattled administrators to be unhelpful and even self-serving. Burger tells how the White House of 1970, anxious to do "something big" about public health but unsure of what, was disconcerted by a PSAC panel's view that a rapid growth of the force of practising physicians was unnecessary but that more money should be spent on biomedical research. The question remains unanswered whether a more pointed commentary on the options before the administration might have headed off President Nixon's "war" on cancer.

Administratively, Dr Keyworth's latter-day version of PSAC may be an improvement on the original. Even from the narrow view of the technical community, there is a balance to be struck between the prestige and independence of the old committee and the possibility that its successor may be listened to more attentively. The "bring back PSAC" campaign of the past nine years has been unreflective about this point. And even if, in the months immediately ahead, the chief function of the new science council is to strengthen Dr Keyworth's own position in the White House, that in the long run can do no harm. It may also help that the President's Science Advisor will be the sole channel for communication between the new council and the Administration. He, no doubt, is likely to be best placed to judge when this or that study can most effectively be put into circulation — and which had better be kept in a drawer. The other side of the coin is that the new council will not become a focus for the attention of the technical community, a legitimized whispering tube, and that even the members of the new council may from time to time suspect that they are being used.

The field of operations for the new council is less certain. From some remarks let slip in London last week, Dr Keyworth seems to think of the new council as a means of helping him solve some awkward administrative problems — making different uses of well-established national laboratories for example. But how? A high-level survey of the resources and capabilities of the national laboratories would benefit the whole technical community, even outside the United States, but Dr Keyworth seems to have more particular goals in mind. Does he want the committee to be his hatchet-man? Or, more machiavellian, does he hope that a committee which includes Dr Harold Agnew (once director at Los Alamos) and Dr Edward Teller (closely identified with Livermore) is bound to resolve the ancient dilemma of deciding which laboratory should go? The most obvious danger is that if the committee is too much used as a means of confirming administrative decisions, or as a means of sanctifying prejudices about what the national laboratories can and cannot accomplish, its reputation in the technical community will be unenviable.

Even without a single meeting to its credit, the composition of the new council cannot fail to excite prurient speculation. For the first time in almost a decade, the White House now has a science adviser who is not afraid to confess his enthusiasm for high-energy physics, and many of the members of the new council can be counted on the back of his hand. But should not Dr Keyworth be mixing in wider circles? The appointment of Dr Edward Teller to the council is similarly disconcerting. Nobody disputes Dr Teller's ingenuity or his zeal. That he should be thought of as a hawk is beside the point. But his inconsistency and intellectual impulsiveness are hardly assets for a new committee.



A strategy for coal

The belief that indigenous coal production is strategically essential neglects lost opportunity.

Sir Derek Ezra, the chairman of the National Coal Board, is understandably cross to have been told (*Nature* 21 January) that the industry that he runs with such distinction should go out of business. His letter on page 642 makes a number of points that can readily be conceded. British coal-miners have indeed become more productive; output per man has increased by more than 50 per cent in the past 30 years, consonant with the further increase of three per cent in 1981. Many of the difficulties of the British coal industry can indeed be blamed on backward technology and inadequate investment over a substantial period of time; these were, after all, some of the reasons why the British coal industry was nationalized in 1947. It is also true that the Commission on Energy and the Environment, whose chairman is Lord Flowers, in its report last year supported the understanding between the coal industry and the British government that there should be a long-term programme of investment in British mines; the commission's report, *Coal and the Environment*, would no doubt have been more widely read and thus more influential if it had not been priced out of the market (at £23.00, from Her Majesty's Stationery Office) and then not distributed, on the grounds of cost, to those who might have reviewed it. But neither these arguments nor all the social and thus political arguments that attend coal-mining (not merely in Britain) are sufficient reasons for not asking whether Britain can afford its coal industry.

The central issue is the assumption that not merely coal, but British-mined coal, is strategically necessary now and in the future. The question arose earlier this year with the prospect (now happily averted) that the coal board's final offer to British miners on pay for 1982, widely advertised as the most it could afford, would nevertheless be scorned and some part of British industry once more brought to a halt. This annual tussle between the miners and the board, bodies otherwise entirely in concert in their conviction that public investment in British coal mines must be sustained, has in recent years been emphasized by the discovery that the cost of importing coal into Britain even from Australia is frequently less than that of the home-produced product. Sir Derek Ezra rightly argues that the price of job-lots of coal bought at the depths of a slump are not representative of the prices likely to be charged for it on the open market when prosperity returns again. This point was conceded, but perhaps too grudgingly, in the article now complained of. But who can suppose that if and when international prosperity returns, it will be possible to moderate miners' demands on the size of the British share that should come their way?

None of this implies that British miners are more grasping than miners elsewhere, or that miners as such are a cupiditous lot. The reality is rather that coal-mining is a dirty and dangerous job, especially in deep mines. Many people who do not work in deep mines say they would never do so. Some, perhaps many, who do work there would prefer to work elsewhere, given half a chance. Nobody blames them. Nobody should. So it seems probable that if and when prosperity returns, not merely will cheap job-lots of coal be harder to come by but so will miners — as in the mid-1960s, when Sir Derek's predecessor, Sir Alf (now Lord) Robens laid the foundations of the now conventional wisdom that British coal production at about the present rate (a third down since 1950) is a fixed point in the compass of the future. But mining, however well mechanized, remains labour-intensive, so that the return of prosperity to Britain (there are no mines worth speaking of in Northern Ireland, so that the nomenclature is permissible) will herald either the flight of miners from the mines or their demand to be paid even more above the odds or, more probably, both. To anticipate this is not to slander an honourable group of men but, rather, is to acknowledge their wit and their good sense.

The argument can even be turned around. In the tight

communities of mining Britain, miners are acknowledged to be resourceful as well as stubborn. They do not cut the figures of people who are especially well paid. How could it be otherwise? Even if coal could be sold at the pithead for £100 a tonne (roughly the equivalent of the landed cost of oil in Britain), which is mercifully out of sight, coal-miners would still be producing less wealth than most of those who work in old-fashioned modern industries — motor-car manufacturing, for example. Their capacity individually to create as much wealth as those who work in truly modern industries — the design of microcircuits for example — is certain always to be small. So why not consider what the circumstances would be like if the part of the British work force that rightly prides itself on its special skill was working in a really productive industry? The hidden costs of the conventional British assumption that coal is strategically necessary, reflected not only in subsidies by taxpayers but in needlessly high prices, are certainly high. The opportunity costs of employing such large numbers of talented people on work characterized by its rigours and its hazards hardly seems wise. (The Commission on Energy and the Environment, no doubt because of its terms of reference, does not take up this question.) So why not follow some different strategy — buy fuel on the open market, or buy less labour-intensive ways of generating energy nuclear power stations for example. And none of this is new — W. Stanley Jevons argued, more than a century ago, that the time would come when coal-mining would no longer be competitive in Britain.

Mercury handcuffed

Liberalizing British telecommunications is uphill work because it goes against the grain.

All concerned seem to be making heavy weather of the attempt in the United Kingdom to liberate the telecommunications industry from the monopoly of the British Post Office (now known in this guise as British Telecom). For the best part of two years, the government was struggling with the unfamiliar problems of how in the abstract to define the conditions under which others than the public monopoly might be allowed to compete with it in the market place. Now, with the Telecommunications Act on the statute book, it is possible to buy telephone handsets retail in British shops without breaking the law. The instruments are those to which the monopoly has agreed to recommend that a licence can safely be given, on technical grounds. It has, by all accounts, been less easily persuaded that it should also fall in with the proposals by Cable and Wireless and Barclays Bank to set up an independent trunk telephone system. Called Mercury, the new system was announced last summer as the commercial alternative to British Telecom.

Only this week has an agreement been reached by the Department of Industry, the Mercury consortium and British Telecom. For some unaccountable reason, the parliamentary draughtsman of the bill supposedly ending the monopoly foresaw the possibility that British Telecom's network might at some point be used by communicators other than itself, but failed to provide that those licensed to communicate electromagnetic circuits to each other might use the public network if that were convenient. The public monopoly has therefore been legally within its rights, in the past few weeks, to agree that the alternative network could go ahead but that it would not be allowed to use its own interconnections, domestically (which is less important) and internationally. The trouble, for the monopoly, is that this legally valid position was politically untenable. The Department of Industry could not have lived with such an impasse. So, in the end, all three parties have compromised. British Telecom agrees that Mercury can use its connections, Mercury has agreed to pay a royalty to the monopoly — and the Department of Industry is probably in a mood to proclaim that competition is not the universal boon it is cracked up to be. What it should have done is to have defined the rules of competition and then to have delegated their interpretation to some regulatory body or the courts.

New White House science council

Physics looms large on advisory panel

Washington

Dr Jay Keyworth, President Reagan's Science Advisor, announced last week that he is setting up a new White House Science Council to advise him on "science and technology issues of national concern" and of "changing perspectives in the science and technology communities".

The thirteen members of the new council will carry out tasks similar to those of the President's Science Advisory Committee (PSAC), which was abolished by President Nixon in 1973. However, unlike PSAC, which reported directly to the president, the council will report to Dr Keyworth as director of the Office of Science and Technology Policy.

Dr Keyworth in turn reports to the president's chief-of-staff in the White House, Mr Edwin Meese, implying a lower status to the Science Advisor's job than in previous administrations. So the Science Council, which is also distinct from the commission proposed by Congress in the Science and Technology Policy Act of 1976, will probably enhance Dr Keyworth's position on key policy issues.

The council, already being referred to as WHSC (or "wissick"), will therefore perform many of the functions of the President's Science Advisory Committee, established by President Eisenhower in 1957 in response to the launching of the Russian sputnik, and abolished by President Nixon in 1973 when some of its members took a public stand against his policies on anti-ballistic missiles and supersonic aircraft. But officials at the Office of Science and Technology Policy (OSTP) emphasize that whereas the committee formally reported directly to the president, WHSC will report to Dr Keyworth, OSTP's director.

There was no official comment on the balance of academic interests of the members of the new council, with its heavy emphasis on physics and the "hard sciences", and, apart from Dr Donald Fredrickson, the former director of the National Institutes of Health, no expert from the life or social sciences.

It was stressed, however, that individuals had been selected as much for their broad perspective and judgement on science policy questions as for their particular areas of expertise, and that each member had been hand-picked by Dr Keyworth "after a considerable amount of deliberation".

"We were concerned to have persons

who can deal with the range of issues that we understand to be the greatest interest to this office," said OSTP staff member Dr Thomas H. Johnson, who will act as the executive director of the new council.

The chairman of the council will be Dr Solomon J. Buchsbaum, executive vice-president of Bell Laboratories in New Jersey. Dr Buchsbaum is a plasma physicist by training, and is an experienced governmental science adviser, having acted as chairman of the Defense Science Board under President Ford, and chairman of the Department of Energy's Energy Research Advisory Board under President Carter.

Vice-chairman will be Dr Edward Frieman, also a plasma physicist and a vice-president of the La Jolla-based consulting firm Science Applications Inc. Dr Frieman served as director of the Department of Energy's Office of Energy Research during the second half of the Carter Administration, and before that was deputy director of the Plasma Physics Laboratory at Princeton University.

Like Dr Buchsbaum and several other members of the new council, Dr Frieman has worked extensively as a scientific consultant to the Department of Defense.

Keyworth irked by national labs

One of the first jobs for the White House Science Council will be to hammer out a policy for the national laboratories supported by a federal government. And, according to Dr George A. Keyworth, President Reagan's Science Advisor, the national laboratories must be reorganized so as to provide more support for universities on the one hand and for United States industry on the other.

Dr Keyworth was speaking in London at the annual lunch of the Parliamentary and Scientific Committee, a regular event in the social calendar frequently used by British politicians to make promises or threats to their own scientific community. Dr Keyworth's irritation with the US national laboratories, made conspicuous by the blandness of the remainder of his speech, may cause managers of similar laboratories in Britain to fear that their own paymasters would follow Dr Keyworth's hint.

Dr Keyworth said that the original missions of United States national laboratories are frequently outdated, and that the "many billions of dollars" they receive from the federal government are not all used to "further fundamental knowledge or to address technologies pertinent to our national goals". He said that the laboratories have been "subject only superficially to external review" and that their missions had only rarely been reexamined.

Yet, according to Dr Keyworth, the national laboratories have a logical role as centres for cooperative research involving

The military connection is also illustrated by the links which several members have with the Department of Energy's Los Alamos National Laboratory in New Mexico, where Dr Keyworth spent most of his professional career as a physicist before going to Washington last May.

Dr Harold Agnew, for example, was a member of the original Manhattan Project team that worked on the atomic bomb at Los Alamos. He joined the laboratory in 1942, became a director in 1970, and left to become president of General Atomic in 1978.

Dr Edward Teller, now emeritus professor of physics at the University of California and widely known as "the father of the hydrogen bomb", also worked at Los Alamos between 1943 and 1946, and acted as assistant director of the laboratory between 1949 and 1953. Dr George Cowan, a senior fellow at the laboratory, is a radiochemist specializing in nuclear reactions who, like Teller, worked at the University of Chicago before moving to Los Alamos in 1949.

According to a formal notice which appeared in last Thursday's edition of the *Federal Register*, the council will have a

universities and could even help to make good shortages of teachers of engineering and computer science. He also wants the national laboratories to help reduce the present "unacceptably high resistance" to the application of research.

For the rest, Dr Keyworth gave his somewhat bemused audience a self-congratulatory account of how research and development, and federal support for basic research in particular, were second only to defence in the protection they had enjoyed in the two Reagan budgets. Nevertheless, he said, the scientific community would have to accept a greater degree of discrimination — and more responsibility for making discriminatory judgements.

On international collaboration, Dr Keyworth said that the United States had come to realize the limitations of what it could accomplish on its own, and the opportunities for collaboration with Europe and Japan. He thought that the next generation of high-energy accelerators would have to be built collaboratively, and that there were already opportunities for broadening the base of thermonuclear research programmes, at present too exclusively dependent on tokamak designs.

Although Dr Keyworth did at one point emphasize that he was not to be mistaken for Mr David Stockman, director of the Office of Management and Budget, he left many of his audience with the impression of a man still preoccupied with the small print of resource allocation. ●

maximum of fifteen members, and will hold regular meetings up to six times a year. Some of the council's discussions may take place in public, since Dr Keyworth has said that he will comply with the requirements of the Federal Advisory Committee Act, which requires federal advisory committees to hold open meetings unless there is a specific reason that the meeting should be closed.

In addition to the members of the council, OSTP is compiling a list of about 100 other outside consultants from the scientific and industrial community who are expected to be called upon on an *ad hoc* basis to carry out specific studies.

Although both the apparent "military-industrial" bias, and the lack of social scientists — and women — have already received a certain amount of comment in the scientific community, reaction to the announcement of the new council has generally been favourable.

Mr William Golden, a New York banker who is treasurer of the American Association for the Advancement of Science, said last week that although the Science Council was "not as good as a presidentially-appointed group would be", it was better than none. He added that the individual members of the council were "first class", and that the new council should help strengthen Dr Keyworth's standing in the White House, since at present he was "not very high on the totem pole".

David Dickson

Council's members

Solomon J. Buchsbaum (chairman). Executive vice-president, Bell Laboratories

Edward Frieman (vice-chairman). Vice-president, Science Applications Inc.

Harold M. Agnew. President, General Atomic Company

John Bardeen. Emeritus Professor of Electrical Engineering and Physics, University of Illinois, Urbana

D. Allan Bromley. Henry Ford II Professor of Physics, Yale University

George A. Cowan. Laboratory Senior Fellow, Los Alamos National Laboratory

Edward E. David. President, Exxon Research and Engineering Company

Donald S. Fredrickson. Fellow-in-residence, National Academy of Sciences

Paul E. Gray. President, Massachusetts Institute of Technology

Robert O. Hunter Jr. President, Western Research Company

Arthur K. Kerman. Director, Center of Theoretical Physics, Massachusetts Institute of Technology

David Packard. Chairman of the Board, Hewlett-Packard Company

Edward Teller. Senior Research Fellow, Hoover Institution, Stanford University.

Turkish civil rights

Trials ahead

Dr Yeter Ögelman, a Turkish specialist in thermoluminescence and a lecturer in physics at Cukurova University, Adana, until her arrest and imprisonment on 16 May 1981, has now been released on bail. Her trial for helping to organize a women's rights group between 1975 and 1977 will, however, be resumed on 10 March. The prosecution is asking for a prison sentence of between 8 and 15 years.

Dr Ögelman's plight was exemplified some weeks ago when an article submitted for publication in *Nature* recorded her change of address and relocation in prison. The article will be published very shortly.

According to the Turkish authorities, Dr Ögelman's arrest and prosecution fall under the terms of article 141 of the Turkish criminal code. This lays down that those administering societies "with the purpose of establishing domination of a social class or overthrowing any of the established basic economic or social orders of the country shall be punished by heavy imprisonment from eight to fifteen years".

The code has already been used to ban the Turkish communist party. However, Dr Ögelman protests that, contrary to the authorities' accusations, neither she nor the "Progressive Women's Organization" was associated with the communist party. She says the women's organization was formed to campaign for women's rights and for improved education for women in Turkey.

Amnesty International has taken up Dr Ögelman's case, and says that her arrest contravenes the European Convention of Human Rights, of which Turkey is a signatory. That convention guarantees the right to freedom of expression and of association with others, although those rights may be prescribed by law "in the interests of national security or public safety". The Turkish authorities have denied that Dr Ögelman's arrest contravenes the convention.

Dr Ögelman was one of more than 170 people brought before the courts on 15 January, and was one of the lucky fifteen released on bail. Most of those on trial were allegedly members of the left-wing school-teachers' association. Teachers' unions were banned nearly ten years ago, while the association was banned after General Evren's coup in 1980.

A number of academics have been arrested since the coup, and Amnesty International is unsure of their fate. A more general worry facing Turkish higher education is the bill announced last November that will further circumscribe university autonomy — for example, by giving the state control of senior appointments. The government's stated intention is to reduce the universities' tendency to act as foci for political disruption and violence.

Philip Campbell

European Space Agency

Peace declared

There is a new cheerfulness among the delegates to the European Space Agency (ESA). One sign of this, at last week's council meeting, is that the member governments have agreed on a resource budget for the next five years. The practical result is that it should now be possible to settle the annual science budget without waiting for unanimous agreement.

The council agreed to maintain the mandatory budget, which covers both science and the agency's basic running costs, at more or less its present level. Hence about 900 million accounting units (£540 million) will be spent over the next five years. But the director-general has promised to shift the balance in favour of science by making savings on the agency's overheads and by diverting interest earned on capital into the science programme. During the next three or four years, the science budget is expected to increase by about three per cent in real terms.

The increase is unlikely to make a substantial difference to the scientific community, which typically has to wait up to ten years for a particular kind of satellite. But the extra money may help ESA out of some of its difficulties.

The council's agreement is nevertheless something of an achievement. The level of the science budget has been hotly disputed for at least the past year, with some member states struggling to maintain their existing commitments and others, in particular France and Germany, arguing forcefully for a substantial increase. At the end of last year, agreement seemed beyond reach, largely because of Germany's wish to spend 20 per cent more on science after 1983. That would have involved all other member states increasing their contributions proportionately. In the event, Germany agreed to the five-year resource level with the proviso that discussions on the level of the science budget start again in 1984.

Last week's council meeting was notable for the formal announcement that Britain has joined the Ariane launcher programme as a fully-fledged member. Until now, Britain has contributed just over 2 per cent of Ariane's development costs through a bilateral agreement with France. The decision to contribute 3.5 per cent of the cost and to enter the programme proper is a recognition of the early promise of Ariane — and also of the diplomatic need to participate in other member projects.

The quarrels of early last year seem thus to have receded. Then the British hankering after telecommunications and the French after the development of launchers polarized discussion of a ten-year plan for the agency proposed by Erik Quistgaard, the new director-general. The agency then had insufficient new applications programmes to fill the gap left

when Ariane and Spacelab end development. During the past year, however, things have changed. Several new programmes have been approved including Ariane IV, the next stage of Ariane development; a Spacelab follow-on programme; L-sat, the heavy telecommunication satellite; and a microgravity programme. Pressure of work has taken precedence over the ten-year plan. But when the delegates next turn to the long-term strategy they may well find that the events of the past year have ironed out some of their earlier differences.

Judy Redfearn

Agricultural Research Council

Cuts tempered

The Agricultural Research Council is having second thoughts about its intended reorganization of agricultural research. The livestock industry may have persuaded the council to cut the Animal Breeding Research Organization at Edinburgh by a half, not four-fifths as originally intended. But the proposal to halve the cost of the Long Ashton Research Station at Bristol is

to stand. The future of the two stations was the topic of a meeting last week between the council and representatives of the laboratories; final decisions will be taken at the end of March.

If the Edinburgh institute is partly reprieved, other council institutes will be on their guard. The council had hoped to save £3 million on its budget for 1983–84 by economies at the two establishments. Now, it will be £1 million short of its targets, and a call has gone to other institutes for self-imposed economies.

With the next financial year upon it, the council seems just now especially anxious to protect its interests in academic research. Last year, central government provided for only a 6 per cent rise on the wages bill but salaries rose by 7.5 per cent, leaving a shortfall of £0.8 million. As the council spends 95 per cent of its budget in its own institutes and units, and two-thirds of that on salaries, research budgets became unacceptably squeezed.

The council expects a more intense squeeze in coming years and fears a shortfall of £3–6 million in 1984–85. As grants to researchers in universities are to be protected and even increased, savings are being sought in institutes.

The council has nevertheless set its face against closing any institute that relies on commissions from the agriculture ministries for more than half its income. Other criteria used to decide which institutes to cut include the level of recent capital expenditure, future needs for facilities, overhead costs and the chances of successfully transferring work.

The proposal to cut the Animal Breeding Research Organization was based largely on internal assessments that had questioned the value of long-term animal breeding experiments. Originally, the council proposed to reduce the organization to only fundamental work, in particular on animal genetics. Representations from the livestock industry, however, seem to have rescued some breeding experiments, and the cattle blood-typing service may also now be retained.

Research on fruit is over-subscribed due to historical accident, according to the council — hence its decision to abandon pomology (apple-work) at Long Ashton and to transfer other work to the East Malling Research Station. Representations from strawberry, plum and mutant-apple growers, however, will now probably ensure some provision for research in these areas.

The council estimates that it will lose £0.5–1 million earned from ministry commissions. Other costs of the closures, including redundancy payments to approximately 220 staff, will initially increase the burden on the council's income. But much will depend on how many staff required to be mobile under the terms of their contracts can be redeployed in other institutes.

Judy Redfearn

Polish universities

Trouble in store

Further clashes between Polish university students and the representatives of the ruling Military Council for National Salvation are feared in the near future, according to a Warsaw Radio broadcast last Sunday. In one of the regular morning briefings on the state of the country, an unidentified speaker said that classes have resumed just when the anniversary of the "so-called March events" is approaching.

The speaker was referring to the expulsion on political grounds of two Warsaw university students, by a decision of the then Minister of Education and Enlightenment, now president, Henryk Jablonski (5 March 1968). Three days later, a mass protest was held in the university grounds, claiming that the expulsion was contrary to university regulations since it had not been made by the rector. The students and some sympathetic academics were forcibly dispersed, but the incident triggered similar protests at universities throughout Poland — and ultimately resulted in mass dismissals and expulsions of students and staff, and the closure of several university departments.

Last year, the March anniversary was marked by a rally of students and academics in Warsaw, and by the formal rehabilitation of the "victims of 1968", including the philosopher, Dr Leszek Kolakowski. Even in the relatively liberal climate of 1981, the commemorations did not take place unopposed; the new "Grunwald Patriotic Association" was set up to hold a counter-meeting to praise government action in 1968.

Last Sunday's broadcast pinpointed growing official fears that young people have not yet accepted the changed situation in Poland and, in particular, the outlawing of the Independent Students Association (NZS). Formally, all protests are attributed to manipulation by hostile elements and political provocateurs. A government leaflet addressed to young people urges them not to be duped into involving the country in a civil war which would be a "tragedy for millions". There are hints in the media of attempts to defuse the tense situation by suggesting that all is not lost. An article in the daily *Zycie Warszawy* on the reopening of the University of Warsaw said that so far, no student had been expelled as a result of martial law — not even those at present interned. At the same time, the danger of taking part in demonstrations is stressed; a Gdansk commentator noted that those students arrested in the demonstration of 30 January would "have to say goodbye to their studies, at least for some considerable time".

Significantly, the announced agenda for the session of the Sejm (parliament), next Friday and Saturday, makes no mention of the new higher education bill.

Vera Rich

Spain back to fold

The European Centre for Nuclear Physics (CERN) near Geneva has been in a pleasant spin for the past few weeks deciding how to react to an advance from the Spanish government. Spain wishes to rejoin (after a brief membership between 1961 and 1968) but there are few people at CERN who remember how such things are done.

In fact the last time a country joined CERN was 21 years ago, when the new adherent was Spain again. But they were early days for CERN, which is now an enormous institution with some 6,000 researchers and staff, massive experimental facilities for elementary particle physics and a £177 million annual budget to match. No official figures are available yet, but accession could cost Spain of the order of £10 million a year, something approaching a quarter of the present Spanish science budget.

The precise price which Spain must pay will be a matter for negotiation. Although members must pay in proportion to their gross national product, and Spain's is moderately high, it might be possible to negotiate step-by-step entry. However, all 12 members of the CERN council must agree to the terms, and there is an argument that Spain should pay entry fee in addition to the annual rate, to account for the fact that it will become part owner of the CERN accelerators. The lawyers are working on that, a spokesman said this week.

Robert Walgate

US student loans

Lean prospects

Washington

Students from across the United States are planning to converge on Washington next Monday for a mass demonstration on Capitol Hill against cuts proposed in federal support for student loans. The demonstration is part of an intense, nationwide lobbying campaign by both students and university administrators aimed at persuading Congress to reject the Reagan Administration's cuts.

The Administration points to the escalating costs of the loan support programme, largely a result of recent increases in US interest rates to record high levels, and is seeking to reduce federal subsidies from an estimated \$3,000 million in 1980 to \$1,600 million by 1983.

Graduate students will be among those heavily affected by the cuts. Under the government's Guaranteed Student Loan (GSL) programme, graduates qualify for loans at a fixed 9 per cent interest rate; under new proposals from the Department of Education, they would now be allowed to borrow only from an auxiliary loan programme at significantly higher costs, with an interest rate of 14 per cent.

The result, according to figures prepared by the American Council on Education (ACE), is that a student who borrows \$8,000 a year on top of previous undergraduate loans to make his or her way through graduate school could be paying off the money at a rate of about \$500 a month for the succeeding ten years.

The presidents of almost all the nation's universities and colleges have loudly denounced the proposed cuts, claiming that by discouraging students from entering graduate training unless their sources of income are secure, the Administration is discriminating against groups who lack this security and undermining the long-term economic health of the nation.

Many feel that they had accepted last year's 20–25 per cent cuts in student loan support as their appropriate share of broader efforts to reduce federal spending. The new cuts, which in some cases would reduce federal subsidies by up to 50 per cent, have provoked the feeling that students are being asked to carry an unfair burden of Mr Reagan's economic policies.

The Reagan Administration's proposals are a mixture of economic pragmatism and ideological conviction that federal subsidies should be confined to the "truly needy". At present, under the GSL programme the government pays the 9 per cent interest on loans for both undergraduate and graduate students while they are at college, and also pays the difference between 9 per cent and the current interest rate for the remaining period of the loan.

Mr Terrel H. Bell, the Education Secretary, said in Washington soon after the budget proposals were announced that,

as a result of soaring interest rates, the cost of administering the guaranteed loans, \$367 million in 1977, is expected to rise to \$3,400 million in 1983.

Reducing the eligibility requirements for the guaranteed loans and shifting the graduate students to the auxiliary loan programme would, claims the Administration, not only ensure that the available loans go to those most in need, but that the total costs to the federal government would be decreased by more than \$2,000 million by 1987.

Many of the leading research universities, however, claim that such cuts would have a devastating impact. Dr Edward J. Bloustein, president of Rutgers University in New Jersey, believes they would eliminate more than \$12 million in aid received by about half of the 4,000 graduate students at his university.

According to figures produced by ACE, about 600,000 graduate students — half of those at present enrolled in US universities — are borrowing money under the GSL programme, in many cases close to the limit of \$5,000 a year. At some of the nation's

leading private research universities, the cost of graduate education, including living and tuition expenses, can be as much as \$13,000 a year and university officials argue that, if they are forced to seek the more expensive loans, many students will either give up or delay their graduate education.

Given the extent of popular support for the student loan programme, the proposed reductions are expected to be strongly opposed by Congress.

In their defence, Administration officials reply that the present loan programme "encourages students to borrow regardless of financial need from their first year at school and can needlessly cause students to amass high levels of debt" which in turn allows "decreased reliance on family savings and student work". They also point to other initiatives, such as the efforts by the National Science Foundation to support graduate students in science and engineering, as proof that the government is prepared to continue support where it feels it to be necessary and appropriate.

David Dickson

Indian Antarctic expedition

Going for good

New Delhi

The Indian government, which sent its first expedition to Antarctica last month, is now actively considering the establishment of a permanent Antarctic base by 1985 or even earlier. Until then, the government intends to operate an unmanned weather station that will be linked via the polar Sun-synchronous satellite that India hopes to

manned station this year. The Antarctic programme is also considered essential to raise India's prestige among littoral states and to give it more weight in international forums on Indian Ocean affairs.

Although many of the Antarctic treaty powers have welcomed India to their club, India does not intend to sign the treaty immediately. It is, however, likely to stake territorial claims if the treaty members do not adopt the common heritage principle.

The Indian expedition was not advertised in advance for fear of adverse



launch next year. The automatic solar powered weather station left behind by the recent expedition team will function for six months. It will be replaced in October, when it is hoped that a second team will arrive for a three-month stay.

Plans are also under way to set up an Antarctic Research Centre in India for training manpower, developing research programmes and working out the logistics of base support with the help of air force planes. The Department of Ocean Development is also planning to buy a ship for Antarctic work. One vessel under consideration is the 3,600 tonne *HMS Endurance*, the British Antarctic patrol ship, now for sale at a price of £1.5 million.

There are several reasons for India's new-found interest in Antarctica. Politically India does not want to be left behind China, which is planning to set up a

reactions from Antarctic treaty members. But, as it turned out, the reaction was quite favourable. The Indian team on board the chartered Norwegian ship *Polarsirkel* not only received weather reports from bases in Antarctica but also navigational help from the Japanese icebreaker *Fuji*.

The 21-member team led by Dr S. Z. Qasim landed on 9 January at a location 70° 3'S and 40° 7'E — an abandoned Soviet campsite. The Indians named the place "Dakshin (South) Gangotri". (Gangotri is the source of the River Ganges in the Himalayas.) The team left Antarctica after camping for 10 days, bringing back a tonne of rocks, several samples of ice and aerosols. One notable observation by the team was an unrecorded undersea mount in the Indian Ocean which has now been named Mt Indira after the Indian Prime Minister.

K. S. Jayaraman

Acid rain

Swedes persist

Brussels

Sweden is yet again to hold a major ministerial conference in an attempt to persuade its European neighbours to take the country's acid rain problem seriously. In June a two-week conference is to be held in Stockholm attended by 15 environment ministers from East and West — the sole country not being represented at a ministerial level being the United Kingdom. The first week of the conference will be devoted to examining the scientific evidence, and a political debate will follow.

It will be hard to separate science from politics. Many of the political arguments rely on demonstrating convincingly the direct link between sulphur dioxide emissions from the coal and oil burning industrial centres in Europe and the increase in the acidity and the presence of heavy metals in surface water in Scandinavia.

At a recent meeting in Brussels, organized by the European Environmental Bureau, many experts admitted that nobody was entirely clear about the precise

year, well above that from West European states. The poor economic situation and energy problems of Comecon countries make it unlikely that they would consider investing money in costly anti-pollution equipment.

The 1979 Geneva Convention on long-range transboundary air pollution merely binds signatory states to "using the best available technology which is economically feasible". The Geneva Convention is still not in force, although 34 countries have initialled it. In the European Community, Greece, Belgium and Italy have so far failed to ratify.

Sweden's attempts to create the political momentum needed for change include the UN conference in Stockholm in 1972, the Helsinki meeting in 1975 and last year's acid rain conference in Gothenburg, but these initiatives have so far produced few concrete actions. Despite the ammunition provided by a study by the Organization for Economic Cooperation and Development, the European Commission has also remained far from convinced that there is a need to do anything more than monitor the situation. Many member states, like the United Kingdom, remain openly sceptical about the cause and effect of acid rain.

Jasper Becker

Dutch elm disease

Northern drift

The British battle against Dutch elm disease has been virtually lost in the south of England and Wales, while the chances of success in northern Britain are slim, according to last week's annual review of the state of the disease by the Forestry Commission. In the south, perhaps 20 million of an estimated 29 million elm trees have been killed. And in the north, where beetles of the Scolytidae family have already crossed the north-south divide, the fight against the fungus carried by the beetle seems likely to be hampered by straitened local authority budgets.

The battle strategy recommended for the north of England is one which has failed in the south: sanitation felling. The objective is to cut down trees infected with the fungus *Ceratocystis ulmi* at an early stage and to burn the bark. The Forestry Commission holds out little hope for alternative control strategies now being tested — the "tree trap" technique that uses cacodylic acid (a tree-killing herbicide) to attract the beetles, artificial pheromones in conjunction with sticky traps and the fungicide Ceratotec, which can cure lightly infected elms. These techniques, the commission says, are only supplementary to sanitation felling, and are expensive.

So the labour-intensive technique of sanitation felling is recommended to continue in lightly affected areas. But local authorities must decide on what scale to battle. While control over the movement of elm wood is maintained by the Forestry

Commission, the decision to divert funds to a felling campaign rests entirely with local authorities.

Where the elm population is isolated, active sanitation has proved highly effective. East Sussex County Council began its campaign in 1973, when signs of the disease first showed. Losses there have declined since 1979 and the worst may well be over. In Brighton, the casualty rate is down to 10 per cent. An important factor in the success of the campaign here was luck. The region is isolated and a good distance from the sources of infection — the docks of London, Bristol and the south. The council sees no reason to give up its campaign, and its budget for 1982 — £97,000 — will also cover planting costs.

East Sussex, however, is only a small pocket on a map of devastation. Authorities in the western counties have long since given up the fight and the area, with its dead elms and broken stumps, is described as looking like the Somme 65 years ago. The area to the north of Merseyside and the Wash, where conditions and species of elm are different, is described as "lightly affected", with as few as one per cent of elms lost each year. But the 5 million elms in the north of England, mostly wych elm (*Ulmus glabra*) have an important amenity value and constitute 10 per cent of total northern tree population. A majority of local authorities in the border regions of Scotland and northern England have ceased to exercise their powers of direct felling, some as long ago as 1977.

The problem is twofold. Once the disease takes hold, it is impossible to eradicate. Moreover, the elm bark beetle is no respecter of regional boundaries and so far local authority programmes have been piecemeal. But the cost of sanitation felling falls entirely on local authority funds. Until 1979, the Countryside Commission for England and Wales supported local programmes, but has now been told that there is no prospect of diverting any substantial part of its funds to assist felling campaigns.

Support from central government has been equally unforthcoming. The Department of the Environment has its eyes set firmly over the horizon and describes the present problem as being on a "back burner", pointing to the Tree Council as the body maintaining a watching brief. The Tree Council, however, has been wholly independent, financially as well as administratively, since 1978. Hope for the future, as far as the Department of the Environment is concerned, lies in the development of elms resistant to the disease. Japanese elms are being planted in the royal parks.

The irony is that in seeking to avoid present costs, local authorities may well incur higher future costs. Merseyside County Council is having to spend nearly £250,000 to fell dead and dying elms that pose a danger to public safety. **Jane Wynn**



mechanism between the cause and effect of acid rain. The evidence was there, though, to show that dramatic changes have taken place in the forests and lakes of Scandinavia even in the past two years. Research is now revealing evidence that sulphur dioxide emissions may also be linked to acidity in fresh water in Belgium and damage to forests in Germany.

While much of continental Europe is far richer in limestone than Scandinavia and is, therefore, less vulnerable to increases in the acidity of fresh water, the effect of the acid precipitation is nonetheless becoming apparent. The problem, the environmentalists argue is not just a Scandinavian one, and is already too serious to allow scientists the luxury of building up irrefutable evidence about the mechanisms involved before action is taken.

The political and economic opposition facing the Swedes seem unsurmountable. Eastern Europe is the source of much of the pollution — notably East Germany and Czechoslovakia. Each provides per capita sulphur emissions of more than 100 kg per

CORRESPONDENCE

Coal as Britain's security

From the Chairman of the National Coal Board

SIR — Your uncouthly superficial and misleading article (21 January 1982, p.177) accuses the National Coal Board and the Government of being "in collusion to make water run uphill".

What, in fact, the NCB, the trade unions and the Government are jointly involved in is to ensure the security of future energy supplies for Britain by developing new and reconstructed collieries so that sufficient coal continues to flow to power stations, industry and homes and the country can continue, progressively, to diminish its dependence on oil.

Throughout the 1960s the coal industry was starved of capital when the policy of successive governments was to rely on plentiful supplies of cheap Middle East oil. Since the Middle East crisis of 1973–74 which changed that shortsighted strategy, investment in new and replacement mining capacity — essential to ensure continuity of production in an extractive industry like coalmining — has totalled £3 billion and continues at the level of more than £700 million per year. Because it takes about 10 years to develop a major new colliery, there is an obvious short-term financial burden of meeting heavy interest charges before coal production begins. That is one of the main problems the coal industry faces today and it is the legacy of earlier years of strategic neglect.

Your article scarcely recognizes the industry's considerable achievements. The coal we supply today is about two-thirds the price of oil and meets nearly 40 per cent of Britain's entire energy needs. Productivity in 1981–82 has so far risen by 5 per cent at the coal face and more than 3 per cent overall. After a previous period of decline, output at long-life collieries has increased by more than 7 million tonnes in the past three years. Exports have doubled for the third year running, and will reach about 10 million tonnes of coal and coke, which — with the addition of machinery exports by the manufacturing companies and mining consultancy services — will earn for Britain over £500 million in foreign currency.

Particularly in these times of continuing recession the British economy benefits from major purchases of goods and services from the National Coal Board — totalling £983 million in 1980–81 and having an overseas content of only £26 million (2.6 per cent).

It is astonishing to find you now repeating the same question that was asked in the 1960s of whether Britain needs a coal industry. An independent answer was recently given by the report of the Commission on Energy and the Environment which supported the coal industry's continuing large-scale investment. The Commission, whose chairman was Lord Flowers, stated that: "In both the energy and the environmental interest, we consider that greater stability of long-term planning is the essential pre-condition of successful modernization of the industry".

The Commission's conclusions on coal imports will also interest your readers: "We believe that most western governments will

continue to find it sensible to favour indigenous coal supplies, and to be unwilling to jeopardize sensible long-run developments by allowing customers to play the market by allowing short bursts of low-cost imports of coal during temporary recessions. Most importantly, we recognize the social unacceptability of repeating in the last decades of the century the sequence of creation and then destruction of mining communities."

DEREK EZRA

*National Coal Board,
London SW1, UK*

Commercial risk

SIR — There is surprisingly little discussion about the ways in which the practice of the biological sciences is changing as a result of the massive commercial investment in biotechnology. While some, like myself, may find the growing commercialization of bioscience depressing, there seems little one can do to stop its spread. But there is one aspect of the process that is an outrage, and bioscientists and the public agencies that fund them should unite to prevent its continuing.

Increasing numbers of scientists are giving their monoclonal-antibody-producing hybridoma cell lines to commercial companies and are then refusing to provide the antibodies and/or the cells to their colleagues. It can be argued that by relieving scientists of the time-consuming task of sending materials around the world, commercial distribution aids rather than hinders the distribution process. But in many cases the cell lines themselves are not made available and the cost of the antibodies is prohibitive if they are needed in large amounts.

The withholding of cells and antibodies for financial gain (or even worse, for selfish scientific advantage) is especially outrageous when these have been developed using public funds. Since it is the granting agencies that must ultimately pay for reagents that would otherwise be free and must endure the slowed pace of scientific advance resulting from the restricted exchange of materials, they should be leading the fight against such anti-science practices. Instead, it has been left to a few enlightened journals, such as *Cell* and *Immunogenetics*, to take up the cause and make free availability of cell (and DNA) clones and antibodies a condition of publication. Is it not time that the granting agencies followed suit and made free availability a condition of financial support?

The American Type Culture Collection was set up to provide a free cell banking and distribution service and has recently taken over the cell bank previously located at the Salk Institute, which contains a number of monoclonal-antibody-producing hybridomas. It would be helpful if this admirable facility were extended to include the distribution of those DNA clones and antibodies that are in

especially heavy demand. There could then be no legitimate excuse for scientists not making their clones and antibodies freely available to their colleagues.

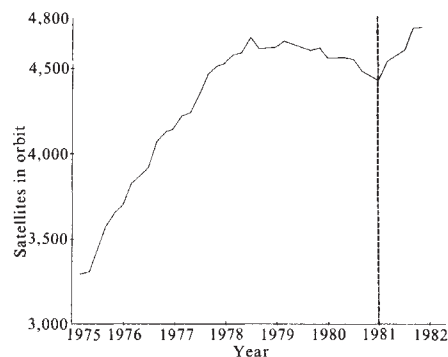
Free exchange between scientists is too important to be given up without a struggle. Without it, not only is efficiency sacrificed, but much of the joy of science will be lost as well.

MARTIN RAFF

*Zoology Department,
University College London,
London, UK*

Spacecraft in orbit

SIR — The suggestion put forward by David Hughes¹ that the number of objects in near-Earth orbit has recently been decreasing, thus lowering the probability of spacecraft collisions, is seriously outdated. The data he presents, which stop at the end of 1980, show a small decrease in the number of objects in orbit from a peak in early 1978 until the end of 1980. However, during 1981 the number of objects in orbit rose rapidly, to an all-time high of 4,740 objects in October 1981 (NASA Satellite Situation Reports). From December 1980 to October 1981 the net increase of objects in orbit was 319. Most of these new objects resulted from two events. Early in 1981, 137 new objects associated with the US Landsat 3 satellite, launched in 1978, were



Data from NASA satellite situation reports. The previous chart¹ finished where the dashed line indicates.

detected. If this is another case of an exploding rocket motor, it indicates that the problem may not yet be solved. Another 118 objects associated with the USSR's Cosmos 1275 were identified shortly after its 4 June 1981 launch. If this rate of increase of objects in orbit continues, the first collision between satellites would be expected in the next 10–15 years according to calculations by Kessler and Cour-Palais². A reversal of this trend is required to prevent a serious hazard to orbiting satellites in the twenty-first century.

GEORGE J. FLYNN

*Washington University,
St Louis,
Missouri, USA*

- Hughes, D.W. *Nature* **295**, 100 (1982).
- Kessler, D.J. & Cour-Palais, B.G. *J. geophys. Res.* **86**, 2637 (1978).

NEWS AND VIEWS

Whence so much ^{15}N ?*from John F. Kerridge*

AN intriguing puzzle is emerging in cosmochemistry about the isotopic composition of nitrogen in the mixture of gas and dust from which the Solar System formed. Recent analyses of meteorite and lunar samples reveal enormous variations in the ratio of the stable nitrogen isotopes, ^{15}N and ^{14}N . The question at issue is whether the variations are due to nuclear processes in the Solar System, mass-dependent isotopic fractionation or mixing of 'solar' nitrogen with one or more alien components — material that has experienced different nucleosynthetic processing from the bulk of Solar System matter. Addition of alien material at levels up to about 5% is already well established from meteorite analyses¹; but the nitrogen data, if the result of nucleosynthetic heterogeneity, could imply much greater levels of alien additions.

Observational data bearing on this issue are numerous and come from a wide variety of sources; a comprehensive review by Geiss and Bochsler will be published shortly². I shall focus here on models advanced to account for those data, addressing specifically the value of protosolar $^{15}\text{N}/^{14}\text{N}$ inherent in each model. Isotopic components will be identified by their percentage depletion or enrichment in ^{15}N relative to the terrestrial atmospheric standard value, $^{15}\text{N}/^{14}\text{N} = 3.66 \times 10^{-3}$.

Geiss and Bochsler² propose that the dominant nitrogen component in the protosolar nebula had a value of +12%; virtually all natural samples could then be explained by admixture of a minor ^{15}N -depleted alien component, possibly pure ^{14}N . Occasional meteorites, such as Renazzo and Al Rais, with values up to +17%, are interpreted as containing interstellar molecules enriched in ^{15}N by ion-molecule reactions at temperatures below 100K in molecular clouds. Such reactions have been identified as the causes of the enormous fractionations observed for the hydrogen isotopes in interstellar molecular clouds, but little corresponding astrophysical evidence exists for nitrogen and it

is debatable whether adequate ionization of nitrogen-bearing species occurs in the interstellar medium. The secular increase from <-21% to +12% observed in nitrogen trapped during the past 3–4 Gyr in lunar surface samples is ascribed to mixing of solar wind characterized by a constant nitrogen value of +12% with an outgassed lunar component rich in alien ^{14}N , whose mixing ratio decreased with time. There is strong evidence from the analysis of lunar samples, however, against the presence of such a hypothetical non-solar wind nitrogen component, and, furthermore, indigenous lunar nitrogen appears to have a composition close to zero rather than <-21% as predicted.

A recent analysis of nitrogen in the Allende carbonaceous meteorite, by Thiemens and Clayton³, has been interpreted in terms of a major protosolar value of -3% and a minor alien component of -9%. A -3% component is also observed in meteorites known as enstatite chondrites, which many believe to contain nitrogen condensed directly from the nebula at relatively high temperatures with minimal isotopic fractionation. It must be stressed, however, that no host phase for nitrogen in any meteorite has been firmly identified with relict nebular material, so the possibility of secondary fractionation is always present. The alien nature of the -9% component is supported by correlations with an alien xenon component recently identified by Frick and Pepin⁴. Thiemens and Clayton interpret meteoritic values up to about +4% as due to mass fractionation during nebular condensation, and the <-21% lunar value as due to fractionation of the -3% component in an early solar wind. The apparent secular increase recorded by the lunar surface samples is ascribed to nuclear reactions, that is, production of ^{15}N , in the

solar wind reservoir. Problems concerning production of ^{15}N in the Sun will be discussed later; specific difficulties of the Thiemens and Clayton model are its failure to account for the +17% meteoritic values and the lack of evidence that mass fractionation in the solar wind could reduce a value of -3% to <-21%.

My own model⁵ of this process which yields a protosolar $^{15}\text{N}/^{14}\text{N}$ of <-21%, assumes that the earliest apparent solar wind value characterized the solar convective zone at that time, thereby establishing an upper limit to the protosolar value. The apparent secular increase in the solar wind value is ascribed to nuclear production of ^{15}N in surface regions of the Sun; meteoritic and planetary values are attributed to mass fractionation during formation of condensed species at very low temperatures, though addition of ^{15}N -enriched alien material is a reasonable alternative. The main difficulty with this model, and the preceding one, stems from the considerable body of observational evidence constraining nuclear production of ^{15}N in the Sun.

The present γ -ray flux from the Sun indicates a level of nuclear activity far too low to generate adequate ^{15}N ; the lack of evidence for significantly greater solar flare activity in the past similarly precludes greater ^{15}N production then; and certain other nuclides, whose production would accompany that of ^{15}N , are not observed. Another problem concerns the temperatures at which adequate isotopic fractionation could occur to explain the meteoritic and planetary data; at such low temperatures reaction or exchange between neutral species, as customarily invoked for nebular condensation, is implausible. Ion-molecule reactions can again be invoked, but with the same reservations as discussed earlier. As an alternative to fractionation, addition of alien ^{15}N -rich matter cannot be ruled out but the scarcity of ^{15}N in the interstellar medium reduces the plausibility of such a component.

John F. Kerridge is in the Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024.

Finally, there is a class of models based, like the previous one, on solar wind data, but which ascribes the apparent secular change to 'contamination' of the Sun's convective zone by interstellar matter. The most detailed of these models⁶ proposes an early addition of material from a planetary nebula rich in ^{14}N to a proto-Solar System consisting initially of +17% material. This would have resulted in a value of <-21% for the solar convective zone which has subsequently evolved towards the bulk solar value of +17% as underlying material replaced ejected solar wind. Meteoritic and planetary compositions may be interpreted as in the Geiss and Bochsler model. A general problem with 'contamination' models is the large apparent size of the actual convective zone; this would require massive additions of alien material for which evidence should be, but is not, found elsewhere.

A potential weakness of models relying on solar wind data lies in the identification of solar wind composition with that of the solar convective zone. Persuasive arguments have been advanced by Geiss and others (see, for example, ref. 7) that such equivalence does exist, but recent spacecraft measurements of solar flare composition suggest otherwise. Mewaldt and co-workers⁸ have reported an isotopic composition of solar flare neon which is distinctly different from that in the solar wind, and they also argue against mass fractionation in the flare as an explanation of the difference. Arguments against fractionation in the solar wind are equally strong, but the flare measurements are at least a warning against identifying solar wind composition with that in the convective zone.

Clearly, an unequivocal estimate of protosolar $^{15}\text{N}/^{14}\text{N}$ is not yet in sight. Indeed, the observations discussed above may signify isotopic heterogeneity for this element in nebulae. It is equally clear that the apparent secular change in solar wind $^{15}\text{N}/^{14}\text{N}$ lacks a satisfactory explanation. The issue of protosolar $^{15}\text{N}/^{14}\text{N}$ may soon be resolved by the Galileo mission to Jupiter. No net fractionation should have accompanied accretion of nitrogen by Jupiter, and all nitrogen should now be present as ammonia; this simplifies the interpretation of the mass spectrometric data that will be gathered. It is to be hoped that the mission will not succumb to the budget-cutting zeal of the present US administration. If not Galileo, who? If not now, when? □

The Vredefort structure still not understood

from Richard A.F. Grieve

THE recognition that impact cratering was a fundamental geological process in the evolution of early planetary crusts¹, together with recent suggestions that faunal extinctions at the Cretaceous-Tertiary boundary may be linked to a major impact event², have heightened interest in the terrestrial record of impact cratering. Approximately 100 known terrestrial structures larger than 1 km in diameter are considered to be of impact origin. One criterion for their recognition as such is the occurrence of shatter cones — cone-and-cup structures, with striated surfaces radiating from parasitic horse tail-like half cones on the face of a master cone and varying in size from less than 1 cm to as much as 10 m (see the figure). These cones, as well as microscopic planar deformation lamellae in minerals, solid-state glasses and high-pressure mineral polymorphs, are characteristic of shock metamorphism. They have been reproduced in shock experiments and in nuclear explosions, occur at small terrestrial craters with meteorite fragments, and have also been observed in meteorite and lunar samples. In a recent issue of the *Journal of Geophysical Research*, there are six papers dealing with various aspects of the geology and geochemistry of the Vredefort 'Dome', south of Johannesburg, South Africa, which is considered to be one of the largest terrestrial impact structures. One of the papers, by Simpson³, interprets the occurrence and orientations of some

shatter cones at Vredefort as precluding an impact origin.

The principal morphological elements of Vredefort are a central core of Archaean basement with an 18 km radius, a 17 km wide 'collar' of steeply dipping or overturned sedimentary and volcanic rocks of the Witwatersrand and Transvaal Supergroups, and a 20–30 km wide outer synclinorium containing members of the uppermost unit of the Transvaal Supergroup. If an impact feature, Vredefort would correspond to what is known as a complex structure, where the original excavated cavity has been modified by uplift in the centre, producing the central core, the penetration and overturning of the stratigraphically higher strata of the 'collar' rocks and the downdropping of the outermost regions to give the peripheral synclinorium. Simpson, however, points to two localities which suggest that at least some shatter cones appear to post-date the overturning of the 'collar' rocks. Such an interpretation conflicts with predictions of an impact model, in which shatter-coning occurs during passage of the shock wave, before uplift of the core and overturning of the 'collar' rocks.

At the first locality, about 40 km west of the centre of the structure, shatter cones are developed in both overturned and right-way-up Daspoort quartzite, a unit of the Transvaal Supergroup. The direction and plunge of the cone striae indicate that the master cone axes have nearly equivalent



Shatter cones in carbonate rocks at the Houghton structure on Devon Island in the Canadian Arctic.

1. Begemann, F. *Rep. Prog. Phys.* **43**, 1309 (1980).
2. Geiss, J. & Bochsler, P. *Geochim. cosmochim. Acta* (in the press).
3. Thieme, M.H. & Clayton, R.N. *Earth planet. Sci. Lett.* **55**, 363 (1981).
4. Frick, U. & Pepin, R.O. *Earth planet. Sci. Lett.* **56**, 64 (1981).
5. Kerridge, J.F. *Nature* **283**, 183 (1980).
6. Ray, J. & Heymann, D. *Proc. Conf. Ancient Sun*, 491 (Pergamon, Oxford, 1980).
7. Geiss, J. *Proc. 13th int. Cosmic Ray Conf.*, 3375 (1973).
8. Mewaldt, R.A., Spalding, J.D., Stone, E.C. & Vogt, R.E. *Astrophys. J. Lett.* **231**, L97 (1979).

vertical orientations, regardless of bedding orientation. The overturned and right-way-up beds are separated by a fault, which strikes radially towards the centre. If, as suggested, this fault is related to the uplift of the central core, then the similar orientations of the master cone axes indicate that at least one generation of shatter-coning took place after the overturning of the 'collar' rocks.

At the second locality, shatter cones occur in an indurated fault breccia, which lies 40 km north of the centre in a fault zone that is part of a series of faults which downthrow overturned Daspoort quartzite beds. As there is no evidence of shatter-coned clasts in a non-shatter-coned matrix, shatter-coning took place after breccia formation. Furthermore, the faulting is thought to have occurred simultaneously with or after the overturning of the beds, so the shatter-coning is post overturning. Simpson's interpretations of these shatter-coned localities are at odds with an impact origin for Vredefort and, by implication, they cast doubt on shatter cones and other shock features as criteria for the recognition of impact structures.

What then are the alternatives? Simpson suggests the cones resulted from shock waves produced by internal tectonic processes. But the problem with such an explanation is that experimental data indicate that shock features are formed only at pressures of tens to hundreds of kilobars. Shatter cones, in particular, appear to require pressures of 20–60 kbar⁴. Even if such pressures could be maintained in the Earth's crust, how could they be produced? Mechanisms involving a deep internal explosion fail to generate such pressures, because the mechanical coupling efficiency between shock waves in a gaseous medium and rock is low (probably less than 1 per cent).

Previous measurements at about 50 sites by Manton⁵ indicated that virtually all shatter cones, when restored to their pre-overturning position, point to the centre of Vredefort and define a focus, or point of origin, for the shock wave some 13.5 km above the base of the Witwatersrand. This corresponds roughly to ground zero at the time of cone formation. Perhaps these two new occurrences are exceptions. Rare cases of cones which do not point to the centre are known from other structures and the formation of occasional inverted cones is predicted in one hypothesis of cone formation involving differential shock impedance of strata. The occurrence of cones in a fault breccia and not in the associated quartzite is certainly unusual. Shatter cones are generally best developed in fine-grained, structurally homogeneous rocks such as carbonates and quartzites and not in inhomogeneous rocks such as

breccias. The interpretation of the relative timing of the faulting in the two areas with respect to core uplift is critical. In considering the preservation of coesite and stishovite, high-pressure polymorphs of quartz, in pseudotachylite, Martini⁶ noted that some deformation associated with Vredefort is related not to impact but to earlier and normal tectonic action. The injection of pseudotachylite veins into the fault breccia at the second shatter cone locality could be interpreted to indicate that at least some of the faulting predates the shock event.

Whatever the explanation, the bulk of the shatter cone evidence from Vredefort still appears to favour an impact origin. It has taken some twenty years of documen-

tation, experiment and theory to establish the impact origin of shock-metamorphosed structures and to appreciate the complexities of the impact process. The new evidence from Vredefort is enigmatic. It must be weighed against the wealth of other evidence and, in the present author's opinion, is insufficient to invalidate the interpretation of Vredefort as an impact structure and the use of shatter cones and other shock features as indicators of meteorite impact. □

1. Grieve, R.A.F. & Head, J.W. *Episodes* 2, 3 (1981).
2. Alvarez, L.W. *et al. Science* 280, 1095 (1980).
3. Simpson, C. *J. geophys. Res.* 86, 10701 (1981).
4. Roddy, D.J. & Davis, L.K. *Impact and Explosion Cratering*, 715 (Pergamon, Oxford, 1977).
5. Manton, W.I. *Ann. N.Y. Acad. Sci.* 123, 1017 (1965).
6. Martini, J.E.J. *Nature* 272, 715 (1978).

Carbon dioxide and climate: has a signal been observed yet?

from Starley L. Thompson and Stephen H. Schneider

In a recent paper Kukla and Gavin of the Lamont-Doherty Geological Observatory¹ examine recent and historical data on the extent of polar ice and snow, hoping to detect changes that might be a signal of warming caused by the build-up of CO₂ in the atmosphere. This and other new work which looks not only at the effects of changing carbon dioxide levels but also at other gases capable of creating a 'greenhouse effect', make it necessary to consider seriously whether the question posed in the title can now be answered in the affirmative.

Kukla and Gavin compiled recent data for sea ice in the Southern Hemisphere for the period 1973–1980 from weekly charts prepared using satellite IR, visible and microwave observations, as well as conventional surface observations. They selected the late spring and summer periods for examination, for the energy balance model of Ramanathan *et al.*² suggests that the largest CO₂-induced changes would occur then (though the model was actually based on snow cover rather than sea ice change). They found that the mean extent of the summer ice decreased from 8.2×10^6 km² in 1973 to 5.7×10^6 km² in 1980. However, as sea ice varies naturally to a comparable degree the decrease is not necessarily indicative of CO₂-induced warming. Indeed Kukla and Gavin note that the area of ice increased in 1981.

Several atlases published before 1970 report the extent of Southern Hemisphere sea ice to have been greater than that observed during the 1970s. However, as Kukla and Gavin carefully explain, the historical and the recent sets of data are not necessarily comparable as the extent of the ice recorded in the atlases was taken from relatively few surface observations. Observational biases could exist which

would give the false appearance of a long-term trend.

Arctic data are even more equivocal. Kukla and Gavin do not find that similar observations of Northern Hemisphere sea ice indicate that there was less ice now than in the 1930s. Satellite-derived records of summer snow-cover do not show a decrease either, although recent positive surface temperature anomalies at 'selected latitudes' in the north have been associated with the belt of melting snow in spring and summer, and are thus 'consistent with' the Ramanathan *et al.* model predictions for CO₂ increase.

Kukla and Gavin suggest that "highly variable components of the climate system such as snow and pack ice fields" be given priority in the search for early signs of a CO₂ effect on climate, a conclusion consistent with Manabe and Stouffer's model³, whose sensitivity to CO₂ depends heavily on concomitant sea ice changes. Although high variability of some climatic component may imply high sensitivity to external forcing (for example, increased CO₂), such variability unfortunately also makes it difficult to detect a CO₂ signal — even a large one — because high variability usually implies a high noise level as well. The best 'precursor indicators' are, of course, those having a large signal-to-noise ratio.

Chlorofluorocarbons (CFCs)⁴, methane

Starley L. Thompson is based at the Department of Atmospheric Sciences, AK-40, University of Washington, Seattle, Washington 98195. At present, he is with Stephen H. Schneider* at the National Science Foundation sponsored National Center for Atmospheric Research PO Box 3000, Boulder, Colorado 80307.

*Any opinions, findings and conclusions or recommendations expressed herein are those of the author and do not reflect the views of the National Science Foundation.

Richard A.F. Grieve is a Research Scientist at the Earth Physics Branch, Department of Energy, Mines and Resources, Ottawa, Canada. He is currently a Visiting Professor at Brown University, Providence, Rhode Island 02912.

and nitrous oxide⁵ are, like CO₂, effective absorbers of IR radiation and thus could also help to create a greenhouse effect.

The sum of all trace greenhouse gases has even been referred to as 'virtual CO₂ concentration'⁶. Lacis *et al.*⁷ at NASA's Goddard Institute for Space Studies (GISS) recently used a one-dimensional, radiative-convective model to estimate the global equilibrium and transient warmings expected from trace gas increases between 1970 and 1980. Interestingly, the computed equilibrium warming from CH₄, N₂O and CFCs combined was 0.1 °C, nearly as large as the 0.14 °C warming computed from the CO₂ increase during the same decade. There remains, of course, some uncertainty, even in the purely radiative calculations. For example, the equilibrium sensitivity of the Lacis *et al.* model to a 2 p.p.b. (parts per 10¹²) increase in CFCs was 30 per cent more than that in a calculation published by GISS scientists only two months earlier⁸. The more recent result, +0.65 °C, now agrees more closely with Ramanathan's⁴ calculation of +0.9 °C but the above discrepancies remain to be resolved. By allowing for the transient effects from thermal inertia of the oceans^{9,10}, Lacis *et al.* reduced their calculated equilibrium expected warming during the 1970s by about one-half.

The GISS group results come from a mathematical climate model which has an equilibrium global surface temperature sensitivity of about 3 °C for a doubling of CO₂ concentration. This is within the present 'consensus' estimates (1.5–4.5 °C)¹¹. Some still believe that these estimates are much too high^{12,13}. Idso, in particular, strongly believes that 'natural experiments' yield an 'empirical response function' some orders of magnitude smaller than the consensus models although he may have neglected some categories of energy flows and storages. In our opinion, it has been convincingly shown^{11,14} that the neglect of horizontal heat flux, heat storage and total atmosphere/surface energy budget has caused Idso's low estimates of global, equilibrium sensitivity to a doubling of CO₂. We do, however, agree that tentative model assumptions demand empirical verification, but such comparisons between models and observations must be properly applied.

An example of empirical verification of a climate model is given in a recent study by Hansen *et al.*⁸. A simulation using CO₂ forcing was compared with the historical temperature record to give a 'direct' verification of the CO₂ sensitivity of the model. One difficulty with this approach is distinguishing what is still a small CO₂ signal from background noise. In addition, their model was based on assumed solar and volcanic forcings whose combined effects were matched to the observed temperature record by adjusting empirical coefficients. This 'tuning process' can engender confidence when the results are

statistically significant, but this has yet to be shown. Another direct verification approach would be to compare palaeoclimatic temperature proxy data with the results of a model forced with observations — actually reconstructions — of natural long-term CO₂ variations. Such variations have recently been confirmed by Neftel *et al.*¹⁵ to have occurred over at least the past 40,000 yr. Problems with this approach include estimates of global temperature and CO₂ concentration which are subject to large errors and the necessity of accurately accounting for additional climatic forcing factors such as surface albedo changes.

Empirical verification may also be 'indirect' — model performance may be compared with a surrogate 'geophysical experiment' not involving CO₂ change, such as the annual cycles of solar radiation and temperature^{3,16}. Difficulty remains in evaluating the extent to which a surrogate experiment is applicable to the CO₂ forcing case. Both of these empirical methods of verification need to be, and have been¹¹, used to test the 'consensus models'.

The average standard deviation of annual global temperatures within individual decades over the last century is about 0.1 °C. For this reason the estimate of warming during the 1970s given by Lacis *et al.*, despite being comparable with that observed, is too small to engender real confidence in the model results. However, Lacis *et al.* believe that, by 1990, temperature rise signal from the combined trace greenhouse gases should stand out clearly from the background of natural climatic variability. This belief assumes, of course, that current trends in CO₂ and other trace gas concentrations continue, and that the sensitivity of present climate models to such increases is reasonably

correct. A global warming of 0.2–0.3 °C during the 1980s, as Lacis *et al.* predict, would have a fairly small probability of chance occurrence, although it is not unprecedented. An increase of this magnitude occurred in the 1920s (for example, see Fig. 3 of ref. 8), which presumably was not a CO₂-related event. Furthermore, extrapolations of present trends of trace gas increases past the year 2000, particularly those for CO₂ based on the assumption of large coal use, have been strongly challenged on economic grounds^{17,18}. Thus, because a doubling of CO₂ from massive coal use in the near future is questionable, but a further 25–50 per cent increase from petroleum and natural gas combustion is possible, we feel even more strongly⁹ that emphasis for research on CO₂-induced climatic effects should be on the transient response over the next fifty years.

The problem of estimating the transient climatic response also has a bearing on the selection of potential 'precursor indicators' of a CO₂ warming. Many of these suggested indicators tend to be regional in extent, such as the sea ice changes examined by Kukla and Gavin. This is significant because it has recently become apparent that different regions will probably respond at different rates to the continuous CO₂ build-up. For example, a simulation performed with a detailed coupled atmosphere/ocean model published recently by Bryan *et al.*¹⁹ shows sea surface temperatures in near-polar latitudes responding relatively slowly to a stepfunction CO₂ increase compared with the global mean response. Thus, for the purpose of highlighting the most promising regional precursor indicators it is questionable whether one should put much confidence in regional climatic perturbations derived from steady-state simulations in which a CO₂ increase is held fixed²⁰. A better way of identifying regional climatic changes with potentially good signal-to-noise ratios would be to search for these in the results of detailed time-dependent simulations performed with a coupled atmosphere/ocean model using realistic geography and plausible CO₂ increase scenarios¹⁰.

In summary, then, the recent sea ice observations of Kukla and Gavin and the model versus observed surface temperature trend comparisons of the GISS group are at least consistent with consensus modellers' expectations for a climatic signal from increasing 'greenhouse' gases. We believe, however, that the statistical significance of these results is simply too small and the number of unverified modelling assumptions too large to allow one to proclaim 'detection'. But if Lacis *et al.* are right, then *this* decade, not the often cited year 2000, is when a clear warming signal is due. Verification of model predictions by the real atmosphere, of course, takes time; the recent results strongly suggest that the time is closer at hand than previously believed.

- Kukla, G. & Gavin, J. *Science* **214**, 497 (1981).
- Ramanathan, V., Lian, M.S. & Cess, R.D. *J. geophys. Res.* **84**, 4040 (1979).
- Manabe, S. & Stouffer, R.J. *J. geophys. Res.* **85**, 5529 (1980).
- Ramanathan, V. *Science* **190**, 50 (1975).
- Ellingson, R.G. & Gille, J.C. *J. atmos. Sci.* **35**, 523 (1978).
- Flohn, H. *Possible Climatic Consequences of a Man-Made Global Warming* (International Institute for Applied Systems Analysis, Laxenburg, Austria, 1980).
- Lacis, A., Hansen, J., Lee, P., Mitchell, T. & Lebedeff, S. *Geophys. Res. Lett.* **8**, 1035 (1981).
- Hansen, J. *et al. Science* **213**, 957 (1981).
- Thompson, S.L. & Schneider, S.H. *Nature* **290**, 9 (1981).
- Schneider, S.H. & Thompson, S.L. *J. geophys. Res.* **86**, 3135 (1981).
- The CO₂/Climate Review Panel of the Joint Climate Board/Committee on Atmospheric Sciences, Climate Research Committee *Carbon Dioxide and Climate: A Second Assessment* (National Academy of Sciences, Washington DC, in the press).
- Newell, R.E. & Doplick, T.G. *J. appl. Met.* **18**, 822 (1979).
- Idso, S.B. *New Scientist* **92**, 444 (1981).
- Ramanathan, V. *J. Atmos. Sci.* **38**, 918 (1981).
- Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumburn, R. *Nature* **295**, 220 (1982).
- Schneider, S.H. & Thompson, S.L. *Icarus* **41**, 456 (1980).
- Lovins, A., Lovins, H., Krause, F. & Bach, W. *Least-Cost Energy: Solving the CO₂ Problem* (Brick House, Andover, Massachusetts, in the press).
- Ausubel, J.H. *Climate Change and the Carbon Wealth of Nations*, IASA Working Paper WP-80-75 (1980).
- Bryan, K., Komro, F.G., Manabe, S. & Spelman, M.J. *Science* **215**, 56 (1982).
- An example being the mid-latitude, summer 'dry zone' in the simulations of Manabe, S., Wetherald, R.T. & Stouffer, R.J. *Climatic Change* **3**, 345 (1981).

Evolution of photosynthetic pathways in flowering plants

from Peter D. Moore

THE metabolic and anatomical characteristics of plants that possess either Crassulacean Acid Metabolism (CAM) or the C_4 photosynthetic pathway have long interested physiologists and have even begun to stir the imagination of ecologists¹, but relatively few people have applied themselves to the problems associated with the evolution of these mechanisms, beyond the cataloguing of their taxonomic distribution. Briefly, both CAM and C_4 plants have a 4-carbon atom molecule (oxaloacetate) as the first product of photosynthetic carbon fixation, though in the former this process occurs only in the dark. Both have low carbon dioxide compensation points (the CO_2 concentration below which a plant takes up less CO_2 by photosynthesis than it gives out by respiration), are insensitive to atmospheric oxygen concentration and lack photorespiration, by which glycolate is formed and subsequently degraded, thus releasing carbon dioxide. CAM and C_4 plants differ both structurally and ecologically: whereas C_4 plants have the so-called 'Kranz' (literally 'garland') anatomy with modified vascular bundle sheath cells to which the Calvin-Benson cycle carbon fixation system is spatially limited, CAM plants lack Kranz anatomy and the products of carbon fixation are not moved to specialized sites before liberation and

fixation. Both types of photosynthesis are characteristically more frequent in hot, often (though not necessarily for CAM) arid regions, where the stomatal control possible with these modifications may be one of their greatest strategic advantages.

Wide-ranging surveys of the plant kingdom by physiologists and anatomists have resulted in a very considerable list of species in which the CAM and C_4 strategies have been found. Most are flowering plants, but at least one gymnosperm² and two pteridophytes³ are reported to be CAM plants. The figure displays the taxonomic distribution of recorded CAM and C_4 species of flowering plants in a form that makes clear the presumed evolutionary development of the two systems. Two major points emerge: first, both CAM and the C_4 mechanism are widespread in the plant kingdom; yet, second, neither is found in what are considered to be the most primitive orders, such as the Magnoliales, Laurales and Ranunculales (that is, those orders close to the supposed 'ancestral complex' (A), within which the angiosperms originated).

Perhaps the most primitive orders in which the pathways are found are the Sapindales (with C_4 representatives) and the Rosales (with CAM members). These observations lead to two logical conclusions: that the C_3 system is primitive in the angiosperms^{4,5} and that both CAM and C_4 have originated independently and on many occasions during the evolution of flowering plants⁶.

The polyphyletic origin of both mechanisms leads one to suppose that the genetic modifications involved in the adoption of either strategy are relatively small; indeed, the major enzymes required are already available in C_3 plants. One obvious way of investigating further this aspect of the subject is to seek and study plants which are intermediate between, say, C_3 and C_4 in character, either in wild populations of plants or as a result of breeding experiments in the laboratory. Rathnam⁷ has now collated the information derived from such studies, but finds that natural hybridization has been recorded in only four genera of plants. The laboratory crossing of C_3 and C_4 plants, particularly the work of Björkman with *Atriplex*, has not produced a simple picture, even though individuals with intermediate characters have resulted. Rathnam concludes that the likelihood of introducing the C_4 mechanism to C_3 plants by genetic manipulation is not strong.

Peter D. Moore is a Senior Lecturer in the Department of Plant Sciences, University of London King's College, 68 Half Moon Lane, London SE24 9JF.



The diagram on the left shows the distribution of C_4 species within the orders of the angiosperms. The arrangement of orders follows Stebbins¹², in which they are shown to radiate from a proposed ancestral complex 'A'. Stippled orders contain fewer than 30 species with C_4 characteristics, shaded orders contain more than 30 species. Kranz anatomy is taken here to be an adequate criterion for C_4 designation, which may be questionable⁹. D, dicotyledons; M, monocotyledons. On the right is the distribution of CAM species among the angiosperms. Symbols as above. The data given in both diagrams are derived from refs 13–16. The numbered orders of flowering plants contain C_4 or CAM species: 1, Asterales; 2, Polemoniales; 3, Scrophulariales; 4, Gentianales; 5, Polygalales; 6, Geraniales; 7, Sapindales; 8, Euphorbiales; 9, Passiflorales; 10, Violales; 11, Rhamnales; 12, Rosales; 13, Capparales; 14, Caryophyllales; 15, Piperales; 16, Liliales; 17, Bromeliales; 18, Orchidales; 19, Cyperales; 20, Poales (Gramineae); 21, Lamiales.

Perhaps the most satisfying result to emerge from these studies is Rathnam's own work on the millet grass genus *Panicum* and in particular *P. milioides*⁸. This plant has a higher than expected phosphoenolpyruvate (PEP) carboxylase activity and an increased affinity for CO₂, which suggest C₄ tendencies. It has Kranz anatomy, an intermediate CO₂ compensation point of 16 µl l⁻¹ (40–60 for C₃ and 0–5 for C₄) and is less sensitive to O₂ inhibition of apparent photosynthesis than a C₃ species, but more so than a C₄ species. Rathnam proposes that both photosynthetic pathways are operative within the leaf mesophyll of *P. milioides*, the C₃ system being unconfined and not limited to the bundle sheath cells as in the typical C₄ plant. Although the development of the C₄ pathway is incomplete, it would still provide the plant with the selective advantage of being able to concentrate CO₂ at the bundle sheath and thereby reduce photorespiration. On the basis of these findings, Rathnam proposes that the evolutionary development of Kranz anatomy precedes the metabolic modifications involved in the C₄ pathway, in which case Kranz anatomy is not a reliable indication of the C₄ capacity of a plant⁹.

Fortunately, a resolution of these difficulties is suggested by this hypothesis recently put forward by Cockburn¹⁰, who focuses on the mode of operation of the stomatal guard cell. Here, malate (one of the 4-carbon atom molecules involved in both CAM and C₄ metabolism) is known to play an important part and to accumulate in the guard cell during its opening¹¹. The malate is fixed from CO₂ using PEP carboxylase, the precise method found in both CAM and C₃ plants. According to Cockburn, "all the essentials of CAM are represented in stomatal guard cell metabolism". Becoming a CAM plant *sensu stricto* simply involves the expression of these characteristics not only in the stomatal guard cells, but in other cells as well. The development of true C₄ metabolism would therefore appear to be a somewhat more complex series of

modifications involving the development of Kranz anatomy and the use of malate as a mobile carbon resource.

The polyphyletic origin of both CAM and C₄ is thus easy to explain. The question

that remains is which came first, the vascular sheath garland, or the extension of the guard cell malate system? The answer will probably depend on where one looks for it. □

Dynein in the spindle?

from Jeremy Hyams

MITOSIS typically involves two distinct phases of chromosome separation: the movement of chromosomes to the spindle poles (usually referred to as anaphase A) followed by the separation of the poles due to spindle elongation (anaphase B). Although the mechanisms of mitosis remain unknown, the most attractive models envisage that force for chromosome movement is generated by an active sliding displacement between spindle microtubules coupled to controlled changes in their lengths^{1,2}. Motility based on sliding is now well established for ciliary and flagellar beating, another form of movement involving microtubules. In this case, the active component is the ATPase protein dynein which reversibly interconnects the nine outer doublet microtubules of the ciliary or flagellar axoneme³. Crucial to establishing the role of dynein was the demonstration that demembrated sea urchin sperm flagella would resume normal beating movements when supplied with exogenous ATP⁴.

The development of a similar, simple *in vitro* model for studying spindle function has, however, proved to be a far less tractable problem. Spindles are notoriously labile and, in general, may be isolated only in the presence of stabilizing agents, which, by definition, alter the characteristics of the organelle. The best alternative has been to use whole cells made permeable to added reagents by treatment with detergent. Most studies of this type have used a cultured cell line derived from the rat kangaroo called PtK₁. If extracted at the start of anaphase, PtK₁ chromosomes continue to move in an apparently normal fashion as long as the lysis solution contains ATP⁵. More recent studies of the permeabilized cell model have shown that anaphase A and B are physiologically distinct. Whilst the basis of anaphase A remains uncertain, ionic conditions required to maintain anaphase B closely parallel those which support flagellar beating. Chromosome separation is specific for ATP as an added nucleotide, is inhibited by non-hydrolysable ATP analogues and sulphhydryl poisons and reversibly arrested by vanadate, a potent inhibitor of dynein ATPase activity^{6,7}.

In this issue of *Nature* (p.700) Cande reports further evidence for the dynein-like nature of the anaphase B chromosome motor, on the basis of its sensitivity to erythro-9-[3-(2-hydroxynonyl)]adenine (EHNA) which, like vanadate, specifically inhibits flagellar beating and dynein ATPase activity. This physiological evidence for the involvement of dynein in mitosis is supported by the identification in isolated sea urchin spindles of an ATPase which has a subunit molecular weight and ionic and nucleotide characteristics closely resembling flagellar dynein⁸. Although there have been suggestions based on immunological cross-reactivity that the same molecule is involved in both forms of motility⁹, the likelihood is that, if the spindle dynein really does exist, it is distinct from its flagellar counterpart.

The organization of microtubules in the two situations is, for a start, quite different. In the axoneme, dynein molecules form cross-bridges between microtubules which all arise from the basal body and hence are parallel in their orientation. In the spindle, on the other hand, interaction is between microtubules originating from opposite poles and so having opposed polarities¹⁰. Mutants of the green alga *Chlamydomonas* in which the flagella lack the paired dynein arms divide normally¹¹. Likewise human patients in which these structures are missing due to genetic defects suffer only from symptoms associated with ciliary paralysis and are otherwise normal¹². More directly, however, antiserum raised against dynein from mammalian sperm failed to recognize the corresponding proteins in either cytoplasmic extracts or spindles of mammalian cells¹³.

At this point the reader might be forgiven for concluding that a unified, and, better still, experimentally testable view of mitosis is at last starting to emerge. Other evidence, however, suggests that the situation is just as uncertain as ever. Particularly fascinating is a recent report from Aist and Berns¹⁴ involving a more primitive organism, the fungus *Fusarium solani*. These workers used a laser microbeam to sever the region of the spindle between the retreating chromosomes. Surprisingly, chromosome separation in these nuclei did not cease but rather proceeded three times faster than in the undamaged controls. By contrast,

1. See Moore, P.D. *Nature* **280**, 193 (1979).
2. Kluge, M. & Ting, I.P. *Crassulacean Acid Metabolism* (Springer, Berlin, 1978).
3. Wong, S.C. & Hew, C.S. *Am. Fern J.* **66**, 121 (1976).
4. Evans, L.T. in *Photosynthesis and Photorespiration* (eds Hatch, M.D., Osmond, C.B. & Slatyer, R.O.) 130 (Wiley, New York, 1971).
5. Raghavendra, A.S. *Photosynthetica* **14**, 271 (1980).
6. Smith, B.N. & Robbins, M.J. *Proc. 3rd int. Congr. Photosynthesis* **2**, 1579 (1974).
7. Rathnam, C.K.M. *What's New in Pl. Physiol.* **12** (6), 1 (1980).
8. Rathnam, C.K.M. & Chollet, R. *Progr. Phytochem.* **6**, 1 (1980).
9. Hegde, B.A. & Patil, T.M. *Photosynthetica* **15**, 1 (1981).
10. Cockburn, W. *Plant, Cell Environ.* **4**, 417 (1971).
11. Travis, A.J. & Mansfield, T.A. *New Phytol.* **78**, 541 (1977).
12. Stebbins, G.L. *Flowering Plants — Evolution above the Species Level* (Edward Arnold, London, 1974).
13. Downton, W.J.S. *Photosynthetica* **9**, 96 (1975).
14. Raghavendra, A.S. & Das, V.S.R. *Photosynthetica* **12**, 200 (1978).
15. Szarek, S.R. & Ting, I.P. *Photosynthetica* **11**, 330 (1977).
16. Szarek, S.R. & *Photosynthetica* **13**, 467 (1979).

Jeremy Hyams is in the Department of Botany and Microbiology, University College London, Gower Street, London WC1E 6BT.

irradiation of the cytoplasm adjacent to one of the spindle poles caused the chromosomes to move more slowly. Far from supporting a model in which active sliding between microtubules drives chromosome movement, these data are consistent with the spindle acting as a mechanical governor, maintaining a counterforce which resists the pull of some, as yet unidentified, extranuclear component.

It would be quite wrong to conclude that either of the two diametrically opposed views of spindle function presented here is incorrect. On the contrary, mitosis may well involve several distinct mechanisms which are expressed to different degrees in different organisms¹⁵. The new probes of dynein activity such as vanadate and EHNA may well prove valuable in assessing the contribution of sliding in different cell types. Indeed their value may extend

way beyond the world of mitosis. In this issue of *Nature* (p.701), Beckerle and Porter further report that both reagents inhibit particle transport along microtubules in fish chromatophores. It may well be that dynein will be shown to be as ubiquitous a feature of microtubule-based motility systems as myosin is of actin-based ones.

1. McIntosh *et al.* *Nature* 224, 659 (1969).
2. Margolis *et al.* *Nature* 272, 450 (1978).
3. Satir *et al.* *Cell Motility* 1, 303 (1981).
4. Gibbons & Gibbons *J. Cell Biol.* 54, 75 (1972).
5. Cande *et al.* *Proc. natn. Acad. Sci. U.S.A.* 71, 1559 (1974).
6. Cande & Wolniak *J. Cell Biol.* 79, 573 (1978).
7. Cande *Cell* 28, 15 (1982).
8. Pratt *et al.* *J. Cell Biol.* 86, 738 (1980).
9. Mohri *et al.* *Dev. Growth Differ.* 18, 391 (1976).
10. McDonald *et al.* *J. Cell Biol.* 83, 443 (1979).
11. Huang *et al.* *J. cell. biol. Chem.* 254, 3091 (1979).
12. Afzelius *et al.* *J. Cell Biol.* 66, 225 (1975).
13. Zieve & McIntosh *J. Cell Sci.* 48, 241 (1981).
14. Aist & Berns *J. Cell Biol.* 91, 446 (1981).
15. Pickett-Heaps & Bajer *Cytobios* 19, 171 (1977).

Intracellular membrane traffic

from M.J. Geisow

GEORGE PALADE (Yale University), in opening the 92nd Ciba Foundation Symposium* on membrane recycling, reminded participants of a fundamental problem of cell economics. To grow, cells must maintain fluid membranes at ambient temperature. This requirement is purchased at the expense of high molecular diffusion rates ($1\text{--}60\ \mu\text{m min}^{-1}$) which, unless controlled, would lead to randomization of membrane components. When it is realized that animal cells contain up to 10 distinct membranes with characteristic proteins and lipids and that extensive transport occurs between them (the macrophage recycles its $780\ \mu\text{m}^2$ of membrane in 30 min), the full nature of the problem can be appreciated. Nature's

Nature's general solution appears to have been the development of membrane components which can self-assemble to form micro-domains. Within such domains, lateral diffusion rates are very much less than in bulk membrane. Well known examples are the specialized intercellular junctions and clathrin-coated pits. It was from the basis of this generalization, that cells can maintain anisotropic membranes, that the symposium considered how transport between occurs and by what routes.

Cells constitutively internalize their plasma membranes by endocytosis. Two routes of entry of membrane — through coated pits and uncoated invaginations — lead to coat-free, often irregular vesicular bodies in the peripheral cytoplasm of the cell. To designate this poorly charac-

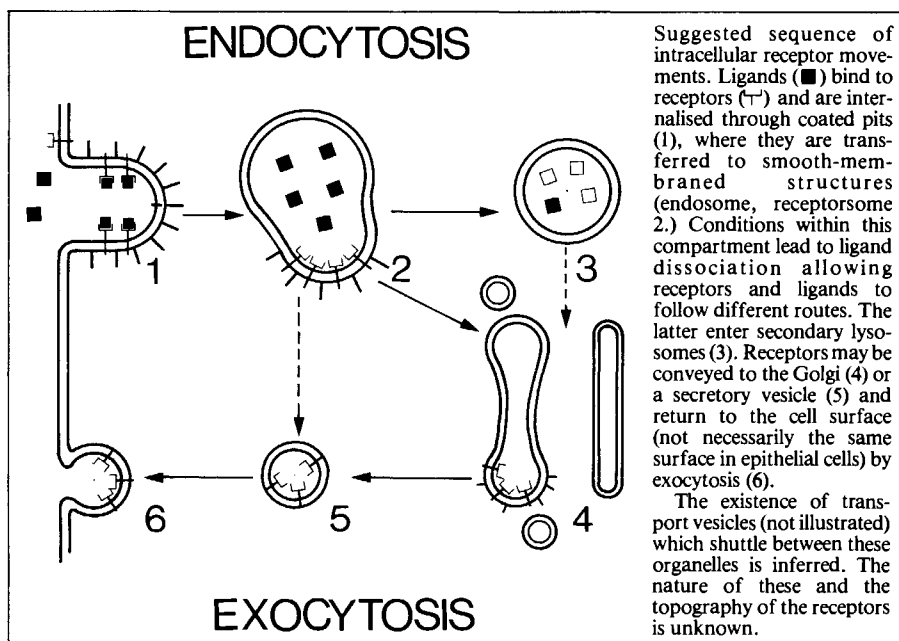
terized compartment, the terms endosome or pre-lysosome have been used, although in the specific case of receptor-mediated uptake initially occurring in coated pits, Ira Pastan and co-workers (NIH) coined the word 'receptosome'¹. It is the ultimate fate of the endosome to fuse with a lysosome.

Clathrin-coated pits are probably microdomains specializing in membrane protein selection. Mark Bretscher (University of Cambridge) showed some time ago that certain cell surface antigens, including Thy-1, are excluded from coated pits². Surface receptors, by contrast, are sometimes present at relatively high surface density in coated

pits. Clustering of receptors for low-density lipoprotein³, α_2 -macroglobulin⁴ and galactose receptor⁵ does not even require the presence of specific ligands and can be explained only by a relatively high affinity of pit for receptor protein. Unfortunately, the basis of this interaction is unknown, because structural analysis of coated membrane has not yet probed the organization within the lipid region under the clathrin cytoskeleton.

Immediately after uptake, receptor and ligand are present in coated vesicles, but within 1–5 min they are in uncoated endosomes. The intermediate coated vesicle, in cultured cells at least, must be a very transient transport element. One good reason for the short lifetime of the coat *in vivo* may be the need for partial or complete uncoating before fusion of the underlying membrane with another organelle. At the Ciba Symposium, James Rothman (Stanford University) described a timely uncoating activity in cytosol, which requires ATP.

Recent work has permitted closer biochemical definition of endosomes. They are not like classical secondary lysosomes, since cytochemistry demonstrates the absence of acid hydrolases. Nevertheless, endosomes may have relatively acidic contents. Spectroscopic studies^{6,7} point to rapid acidification of endocytosed ligands. A clever confirmation of the low pH in the endosome has followed the observation that the fusion of Semliki Forest virions with lipid bilayers only occurs at or below pH 6.0. Viral infection *in vivo* necessitates the fusion of enveloped viruses, taken up via coated pits, with the endosome membrane. Ari Helenius (Yale University) has now shown that this fusion is prevented by weak bases such as amantadine which increase endosome pH (ref.8). Cells can be 'rescued' from irreversible infection only by supplying



M.J. Geisow is in the Biophysics Division at the National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA.

*The Ciba Symposium on 'Membrane Recycling' was held in London on 19–21 January 1982.

the weak base within 4 min of warming cells with adherent virus to 37°C.

The low pH of endosomes may bring about dissociation of ligand and receptor protein. In experiments reported by Anne Hubbard (Johns Hopkins University), liver galactose receptor–ligand complexes are endocytosed at 16°C, a temperature at which endosome–lysosome fusion is impaired⁵. However, ligands still dissociate from receptors and the latter return to the cell surface, despite the absence of ligand degradation. It may be that these endosomes represent the elusive sites of receptor recovery — an acid wash through which regular ‘preening’ of the cell surface can be achieved.

The fate of the ligands and soluble content of endosomes is clearer — they are transferred to secondary lysosomes and degraded, which for them is a one-way trip. Exceptionally, certain receptors, for epidermal growth factor or the Fc portion of immunoglobulin, also enter lysosomes. Whether this is because their affinity for ligands is unaffected by the low pH in the endosome or because of intrinsic differences in the molecular structures of such receptors, remains an open question.

Equally enigmatic is the direction taken by the receptors which do recycle following the dissociation of their ligands or their normal recovery during membrane recycling in the absence of a specific ligand. Specific labels are lost at this stage and anti-receptor immunocytochemistry can only demonstrate the total cellular distribution, not the pathways of recycling. However,

the endocytosis of cationized ferritin (CF), a positively charged, electron-dense marker which binds tightly and non-specifically to membranes, has provided some indirect evidence of how receptors recycle.

Kinetic studies of the endocytosis of CF by pituitary cells and lymphocytes made over a number of years by Marilyn Farquhar^{9,10} show that after its appearance in endosomes, CF enters the trans elements of the Golgi stacks. After 30 min, CF appears on the inner membrane of secretion granules and therefore on the exocytotic leg of its journey through the cell. Autoradiography of iodinated surface membrane proteins suggests essentially the same route — although slow techniques such as conventional fixation and electron microscopy provide images which reflect rate-determining stages in the flux through a cell, so this picture may be incomplete.

In at least one case there is good evidence that CF follows the same route as receptors. Richard Rodewald (University of Virginia) has described transcytosis or diacytosis (passage through a cell) of maternal IgG across the intestinal epithelium of new-born rats¹¹. The immunoglobulin is bound by Fc receptors at the gut lumen and is discharged into the neonatal circulation. Efficient segregation of IgG–Fc receptor complexes from soluble markers occurs in the endosomes near the sites of absorption (coated pits). The integrity of the receptor–ligand complex may be favoured in this case by low pH, since the neutral pH of the neonatal blood affects dissociation of the complex when it appears at the basolateral

cell surface. CF accompanies the ligand through the same pathway and is exocytosed, still in membrane-bound form, at the sites of IgG release.

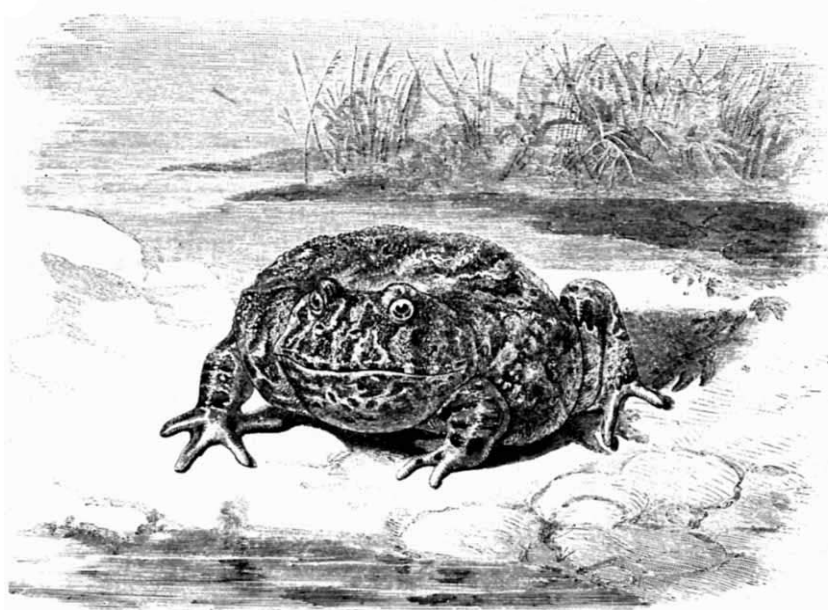
Considerable advances have been made in this popular area of cell biology, to which an increasingly ‘molecular’ approach is being applied. There remain, of course, major problems. We still need to account for the fast recycling times of cell membrane (seconds in the neuromuscular junction) and certain receptors (a minute or less for low-density lipoprotein receptor). Do these membrane components have time to follow the pathways described above? The relationship between endocytosis of membrane components through coated and uncoated regions is also unclear. Most telling of all, however, is the degree of our ignorance of the nature of the signals which direct this extensive traffic of receptor proteins and orchestrate the flow of cellular membrane. To understand these, we need techniques which allow us to observe the membranes from within the cell itself. □

1. Willingham, M.C. & Pastan, I. *Cell* **21**, 67 (1980).
2. Bretscher, M.S., Thomson, J.N. & Pearse, B.M.F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4156 (1980).
3. Goldstein, J.L., Anderson, R.G.W. & Brown, M.S. *Nature* **279**, 679 (1979).
4. Willingham, M.C. & Pastan, I.H. *Cell* **21**, 67 (1980).
5. Wall, D.A. & Hubbard, A.L. *J. Cell Biol.* **90**, 687 (1981).
6. Geisow, M.J., Hart, P., D’Arcy & Young, M.R. *J. Cell Biol.* **89**, 645 (1981).
7. Maxfield, F.R. & Tycko, R. *J. Cell Biol.* **91**, 212a (1981).
8. White, J., Matlin, K. & Helenius, A. *J. Cell Biol.* **89**, 674 (1981).
9. Farquhar, M.G. *J. Cell Biol.* **77**, R35 (1978).
10. Ottensen, P.D., Courtney, P.J. & Farquhar, M.G. *J. exp. Med.* **152**, 1 (1980).
11. Rodewald, R.D. & Abrahamson, D.R. *J. Cell Biol.* **91**, 270 (1981).

100 years ago

Amongst the batrachians exhibited at the Zoological Society are several species of gigantic size when compared with their puny representatives in this country, such as the Agua Toad (*Bufo agua*) of Brazil and the Ocellated Bladder-Frog (*Cystignathus ocellatus*) of Buenos Ayres. But by far the most remarkable of these forms in the series is the Adorned Ceratophrys, or “Esquerzo” of the natives of the Argentine Republic — a large toad of brilliant colours and extraordinary form, of which a figure (Fig 16) is now given, taken from a water-colour sketch prepared by Mr Ernest Griset.

The Esquerzo was discovered by Mr Darwin during the celebrated voyage of H.M.S. *Beagle*, and first described by the late Prof. Bell in the “Zoology of the Voyage of the *Beagle*.” This monster inhabits the pampas of Buenos Ayres, and is said to feed chiefly on its smaller brethren of the same class. Mr Ernest William White, F.Z.S., to whom the Society is indebted for one of the two specimens now in the Gardens, specially mentions it in his lately-published “Cameos from the Silver-Land” as one of the characteristic forms of the grassy plains of the Argentine Republic. “In the damp grass,” he says, may often be perceived the leering eyes and mottled black and green body of the huge Esquerzo (*Ceratophrys ornata*), whose gaping mouth crammed with the body of an unfortunate sapo (toad), and surmounted by threatening horns, inspires terror. This said Esquerzo bears an awfully spiteful character, and is credited with the deaths of many children.



THE ESQUERZO, OR BARKING TOAD

From *Nature* **25**, 393; February 23 (1882).

REVIEW ARTICLE

Very large ground-based telescopes for optical and IR astronomy

J. R. P. Angel

Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

Optical and IR astronomers are taking a hard look at their ground-based facilities and devising new ways of making more economic, bigger and better telescopes. Features of instruments of the 15-m class are likely to include servo control to compensate for atmospheric wavefront errors as well as structural deformation, large honeycomb mirror blanks and mirror surfaces produced by economical techniques developed for aspherics.

WHILE detectors and instrumentation have improved enormously during the past 50 yr, the size of the largest optical telescopes has not increased significantly. Detectors have now reached close to the fundamental noise limit set by photon statistics and techniques have been developed to observe simultaneously many objects in the telescope's field of view. Further progress in many problems of astrophysics now requires an order of magnitude increase in light grasp.

The main purpose of increasing telescope aperture is to reach fainter objects and to improve accuracy of measurement. The quantum nature of light means that to measure intensity, spectrum or polarization to a certain accuracy, even with ideal detectors and optics, requires the collection of a certain minimum amount of energy. For other types of measurements in which detector noise is dominant, the rate of collecting energy is an additional factor, and gives an added advantage to the large telescope. Given that on average only a few hours each day meet all the requirements for observing, and that the largest telescopes must serve a broad community of observers, it is not practical to integrate for more than around 20 h in a season the light from even an object of exceptional importance. For a given collecting aperture this sets a limit to faintness or accuracy.

The qualitative advantages to larger apertures arise from the improved limit to angular resolution set by diffraction. At long IR wavelengths, where image quality is limited by diffraction, high spatial resolution is achieved directly. At shorter wavelengths various interferometric techniques allow the recovery of information on spatial structure at the diffraction limit. This is <10 m arc s in visible light for the telescopes I now consider.

In the US a programme has been started to develop the critical technology for a national telescope with a collecting area equivalent to 15 m diameter. Closely linked projects for smaller telescopes for California and Texas are under way. Meanwhile, in Arizona, the Multiple Mirror Telescope (MMT), a joint research facility of the University of Arizona and the Smithsonian Institution, is a working prototype of many aspects of the new technology. Most ideas for very large telescopes involve multiple primary mirror elements, and the MMT, the first of this type, has six mirrors of 1.8 m diameter all bringing light to a common focus (Fig. 1).

Why ground-based?

The Earth's atmosphere limits the value of ground-based telescopes in various ways. It is completely opaque to wavelengths below $0.3 \mu\text{m}$ while the IR spectrum is broken up by molecular absorption bands, being completely blocked over the range $40\text{--}300 \mu\text{m}$. Background radiation from air scattering of sunlight and moonlight, and from thermal emission in the IR interferes with the observation of faint objects. Refractive index changes caused by thermal gradients in the air prevent one achieving diffraction limited images at shorter wavelengths, and

the uncertainty of weather means one cannot be sure of any particular observing period.

These limitations are removed for telescopes above the atmosphere. The Space Telescope, scheduled for launch in 1985, will overcome many of these problems. However, its aperture of 2.4 m diameter has only $1/40$ the area of a 15-m telescope, while its cost is an order of magnitude larger. Although technology may become so advanced that one could afford very large telescopes to realize the potential of space, our present needs for huge light grasp can only be met by building on the ground.

In fact, many of the observations that need large aperture can be done efficiently through the atmosphere provided that the design, siting and scheduling of the telescope are optimized. Spectroscopy will be the major application for a very large telescope. Provided the atmospheric background is kept to the minimum, the dominant noise source for this work will often be photon shot noise of the signal itself, or detector noise in IR spectroscopy—these noise sources are not reduced by going to space.

Image quality

The factor that most strongly drives cost is the need for excellent image quality. This is required not so much to resolve detail in extended objects but to minimize the background underlying faint unresolved objects. A giant light bucket with indifferent images would be of relatively little value. If, for example, the spectrograph entrance aperture has to be increased in diameter to capture the light of a degraded image, then the increased background may overwhelm the signal. For the faintest objects doubling the telescope diameter is of no advantage if the image diameter is also doubled.

Our intention of achieving good images is likely to have far-reaching consequences for telescope design. Recent experiments and studies demonstrate that some degree of active correction for atmospheric distortion is generally possible, with potentially very large gains for giant telescopes in the IR. If the means to implement such corrections are an integral part of the telescope design, then they may also be used to correct for slight distortion of the optics which would be intolerable otherwise. Large cost savings may be possible if a correction system were to relax the normal tolerance for accuracy, rigidity and stability, especially if the optics are extremely large.

The distortion at the atmosphere can be thought of in terms of the corrugation of a wavefront that was perfectly plane when incident at the top of the atmosphere. It has ripples and waves on all scales, which can be modelled by random turbulence¹. (Note, however, that turbulence inside the telescope housing has only small scales, and causes separate problems discussed separately below.) The random wavefront deviation δ in micrometres between points separated by d across the wave is given by

$$\delta(\mu) \approx (d/d_0)^{5/6} \quad (1)$$

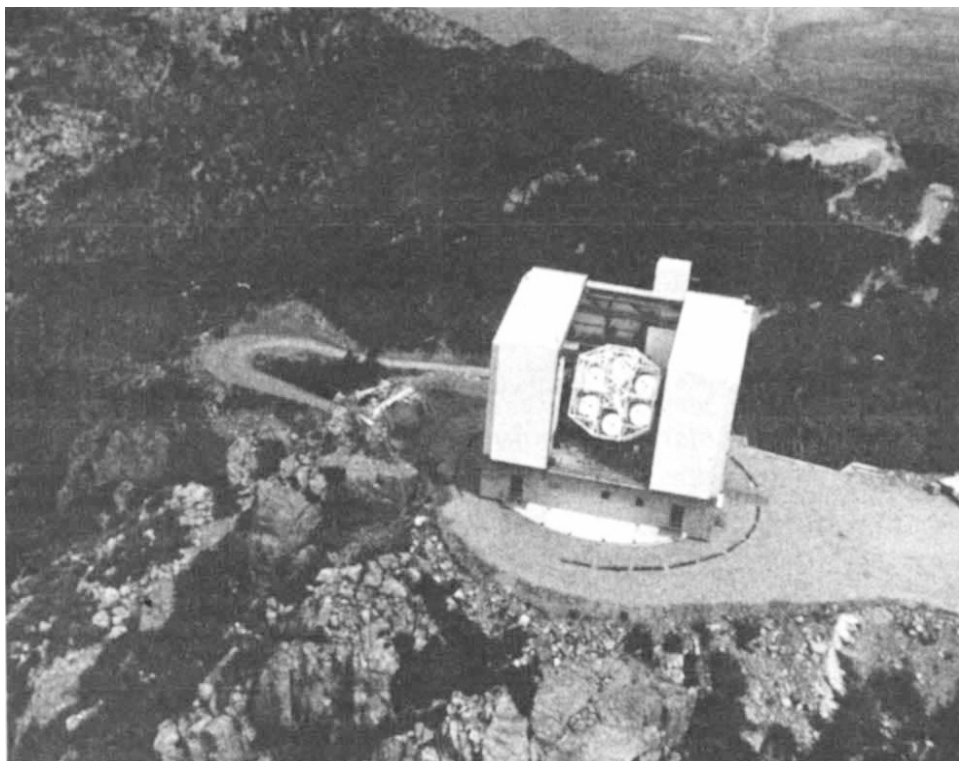


Fig. 1 An aerial view of the Multiple Mirror Telescope in Arizona. The location on top of a sharp, high peak, gives favourable conditions for sharp images. The MMT is jointly owned and operated by the University of Arizona and the Smithsonian Institution.

where d_0 is a measure of the seeing quality and in good conditions may be as large as several metres. The corrugations move at about the wind speed, so that large-amplitude deviations over a scale of 10 m change on a time scale of 1 s, while ripples on a scale of 10 cm vary 100 times faster.

If a true diffraction limited image is to be obtained at any given wavelength, then the deviations across the telescope aperture must not exceed about $\lambda/10$ (see ref. 2). For visible light, the task of correcting the wavefront of a large telescope to this tolerance is extremely difficult. The length scale of ripples with amplitude $>\lambda/10$ is 10–20 cm at best and the time scale is short. Compensation has to be made at an optical surface that can mimic all these rapidly changing small-scale ripples. Considerable progress in making such 'rubber' mirrors has been made by Hardy³. Apart from engineering difficulties, however, a fundamental limitation to achieving complete correction of the visible light wavefront is that of deriving the information needed to make the correction. This requires that the source being observed should be bright, or that there should be a bright source no more than a few arc seconds away whose wavefront, similarly distorted, can be used to derive corrections.

Although full correction of the wavefront is generally not possible, correction of the errors on a scale of metres is. Faint field stars within a couple of arc minutes of a target are adequate to provide the necessary much reduced amount of information. From equation (1) we see that the wavefront tilts δ/d are almost as big at large scales as at small, varying only as $d^{-1/5}$, so significant improvement of visible image quality can be expected. Woolf⁴ has analysed in detail the improvements that can be made in various conditions: typically, 20–30% reduction in image size can be expected in visible light^{4,55}. The better the seeing, the greater the improvement to be made.

In the IR, there is dramatic potential for high quality imaging with large telescopes. At $10\text{ }\mu\text{m}$, for example, the small-scale ripples that do most damage to visible images are below the $\lambda/10$ threshold, and are not noticed. The seeing disk, resulting only from large-scale waves, will be only a little more than half that of visible light. Woolf has pointed out, though, that if the large-scale errors can be measured and continually corrected with the aid of a nearby visible star, fully diffraction limited performance at $10\text{ }\mu\text{m}$ is achieved in reasonably good seeing conditions. For a filled aperture of 15 m diameter the FWHM of the Airy disk is

0.16 arc s. Because almost every source of interest at $10\text{ }\mu\text{m}$ is fainter than sky background, the advantage of operating a large telescope at this high resolution is enormous.

Thermal control

The frequency of occurrence of different degrees of atmospheric turbulence and hence wavefront distortion has been measured with small, low level, test telescopes by Walker⁵. The results show little difference between the good test sites. From these data and balloon observations of turbulence by Barletti *et al.*⁶,

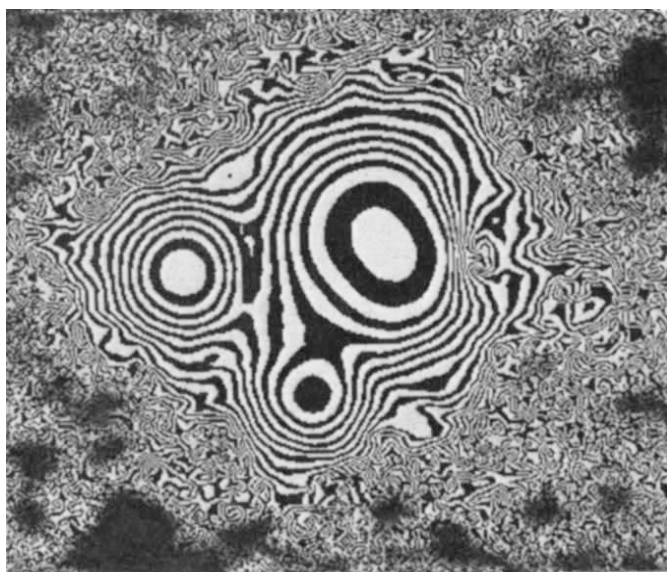


Fig. 2 An example of high resolution imaging from the ground, the QSO PG1115+08. It shows multiple components because of gravitational bending of light rays by an intervening galaxy. The two unresolved round images, separated by 2 arc s, have FWHM of 0.6 arc s. The brightest, elongated image is shown by speckle interferometry to be a double with 0.5 arc s separation⁵⁴. This figure, courtesy of E. K. Hege, was made by Doug Tody at Kitt Peak National Observatory using the IPPS facility with contours space logarithmically by factors of two.

Woolf⁷ has derived probability distributions for long exposure (time averaged) image size. He finds the diameter containing half the energy of a point source should be 0.7 arc s at the 50th percentile, 0.3 arc s at the 10th percentile.

The images that are actually recorded at most large telescopes are rarely as good as this, mostly because of thermal disturbance of the air near the telescope and enclosure. The increased size and thermal inertia of a giant telescope means we must be especially careful to understand and control these local effects. Adiabatic effects of turbulence on refractive index are not significant at normal wind speeds, but heating and cooling by convection and radiation are. The sensitivity to heat can be appreciated when we realize that light passing through a single bubble of air with a temperature difference of only 1 °C from its surroundings is spread into a cone of 0.5 arc s diameter. Turbulent thermal boundary layers can cause acute problems.

Local thermal gradients are potentially worse at high desert sites where diurnal changes of air temperature are large, and where radiation cooling at night is strong. The effects of an observatory building on image quality stem both from heat sources within the building and from failure of external surfaces to accommodate the changes in the ambient temperature. In the MMT building, exposed surfaces were of low thermal mass and were backed by good insulation, and the telescope chamber was small and well ventilated⁸. Beckers⁹ reports a series of measurements at the MMT of temperature made with thermometers and a 10- μ m TV camera, and of internal seeing measured with laser beams in the dome. The use of insulation and surfaces with low emissivity at 10 μ m on the telescope and mount as well as the building, means that very good thermal equilibrium with the ambient air is achieved.

The one part of any telescope that cannot be insulated is the mirror. Seeing generated by mirrors has been studied by Lowne¹⁰, who finds experimentally a seeing disk of 0.5 T arc s for a mirror temperature T °C hotter than ambient. Conventional large solid mirrors have time constants of many hours, and so are generally likely to be out of equilibrium by at least 1 °C from the changing night-time air, even if refrigeration is used during the day to set some anticipated mean operating temperature. Mirror seeing will thus frequently spoil potential sub-arc second seeing, unless the thermal time constant is kept to an hour or less. This goal can be achieved using metal mirrors or glass with faceplates no more than a few centimetres thick and air circulated behind. The fact that the MMT has thin-faced honeycomb mirrors may account in part for its frequently excellent seeing, despite substantial diurnal temperature variation at its high site. The image shown in Fig. 2 of the 'triple' QSO, having high resolution of 0.6 arc s, was obtained with one of the MMT honeycomb mirrors, with correction every 1 s for low frequency image motion caused by large scale wavefront tilts¹¹.

Ideas for very large telescopes

Traditionally telescopes have a single primary mirror made from a rigid massive disk of glass. Mirror supports for optical telescopes are based on the principle that perfect support against gravitational distortion of an elastic solid at any orientation is given by floating it in a liquid of the same density. Mirror cells arrange for the gravity balancing force to be distributed over the mirror back and sides, simulating flotation as far as possible. Thermal distortion of modern mirrors is effectively eliminated through the use of very low expansion materials having coefficient of $<10^{-7}$ per °C. Stiffness is such that the mirror figure is not significantly disturbed by wind in the telescope enclosure.

Mirrors of 4-m diameter are probably the limiting size for an economical telescope using these principles. Considerable increase should be practical if lightweight (thin or honeycomb) blanks are used together with some active compensation for mirror deformation and resultant wavefront distortion. Lewis¹² finds that on site manufacturing of a 10-m blank of 20 cm

thickness out of fused silica or ultra low expansion (ULE) glass is practical. Studies of stiffness and designs for supports of thin meniscus mirrors of ULE or Zerodur in the 7–10 m range have been made by Epps *et al.*¹³, Mack¹⁴ and Pearson¹⁵. Pyrex honeycomb mirrors in this size range that would be considerably stiffer and potentially less expensive are under development^{16,17}. Flotation support alone of large thin mirrors is unlikely to yield a surface precision much better than 1 arc s, particularly when one considers that wind forces in reasonable size domes will be large. However, with the aid of servo controls, as we have considered above for wavefront correction, slight errors in figure may not be a serious drawback.

The goal of a 15-m diameter primary aperture, which I adopt as representative of a very large telescope, is probably too large to be handled practically if made from a single piece of glass. Certainly the technology would have to be proved in smaller sizes before one could consider seriously such an approach. I thus consider designs that involve several or many separate mirror sections. For nearly all optical observations other than interferometry these separate elements could be in the form of separate telescopes, with data from separate instruments added to yield all the signal-to-noise advantage of the full aperture. Disney¹⁸ has pointed out the advantages of such an array, in which fairly high spectroscopic resolution can be obtained with silicon CCD detectors without reaching the detector noise limit. Heterodyne IR interferometry would be possible with such an array¹, but direct interferometry and those IR observations which need all the light at a single focus¹⁹ are almost impossible to achieve if the telescopes are on fixed separate mounts²⁰.

The most obvious route to a single focus is to figure all the elements as parts of a single parabolic surface. These must be mounted in a supporting structure as stiff as possible, and a scheme must be devised to manipulate the segments (or smaller elements reflecting images of the segments) to maintain correct orientation and phase. Such an approach has been studied by astronomers at the University of California. Nelson^{21,22} and his collaborators are developing methods to prepare separately and bring together 25 off-axis hexagonal elements, to make an approximately hexagonal parabolic surface 10 m across. Angel *et al.*²³ have considered a design of six identical circular off-axis mirrors, each 6-m in diameter.

The MMT²⁴ takes an intermediate approach between separate telescopes and a segmented single reflecting surface. The experience of this working prototype has led to proposals⁴ for much larger versions. A design for a 15-m telescope now being studied in some detail incorporates four 7.5-m honeycomb glass primaries co-aligned in a single telescope mount. Many optical observations would be done with relatively small instruments at the four separate foci but, as in the MMT, auxiliary mirrors would be arranged to bring the light to a combined and phase focus when needed. In this concept structural deformation and wavefront errors could be compensated in the small secondary optics or the primary mirrors.

Both the single-surface concept and the multiple telescope concept have the potential of yielding a telescope which can optimize images yet is versatile enough to undertake optical, thermal IR or interferometric observations. These approaches are being considered for the United States national 15-m telescope by a consortium of the Universities of Arizona, California and Texas with Kitt Peak National Observatory. This effort is concentrating on the key areas of technology that need development, bearing in mind a target cost of around \$100 million.

Other concepts for very large telescopes which have received considerable attention are generally less versatile, but may have an important part to play for special tasks. Designs based on the Arecibo radio telescope with a fixed spherical bowl and moving secondary optics^{25–28}, solve the problems of gravitation distortion of a moving primary surface. Spherical segments of the surface could be produced inexpensively and spherical figure can be maintained easily against thermal deformation by laser metrology from the centre of curvature. However, performance

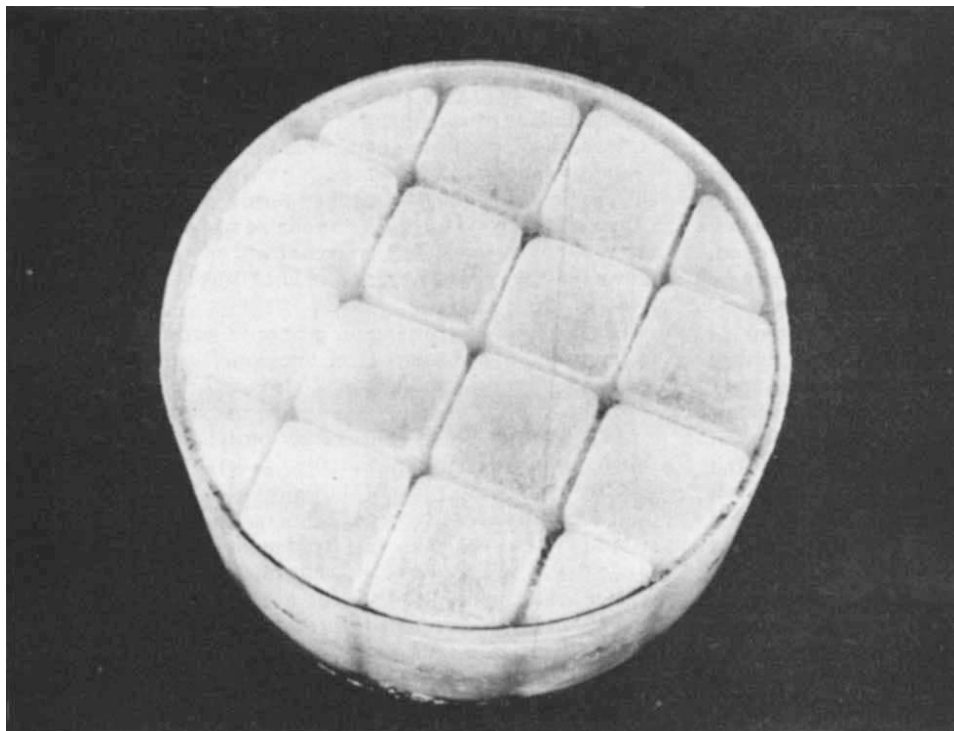


Fig. 3 A honeycomb-faceplate sandwich mirror blank cast from borosilicate glass as a single piece¹⁴: it is 60 cm in diameter and 35 cm thick, with ribs 6 mm thick on 15-cm squares.

in the thermal IR would be poor because of varying background as the detector tracks across different parts of the sphere. Also sky coverage is limited and it is difficult to correct spherical aberration over a large field of view. Furthermore, it would be expensive to raise such a telescope above the turbulent thermal boundary of the ground. A valuable application for this type of telescope would be for multiple object spectroscopy, in which a small corrector would be used for each point object under study, and many such correctors would be tracked at once across a limited spherical focal surface of some 20–30° (ref. 29). Multiple object spectroscopy is, of course, an important application of any of the giant telescopes discussed here.

A large telescope configuration that might become very important, with more experience with interferometry, has been suggested by Low³⁰. Several telescopes would move on circular tracks, so as to remain in a line perpendicular to the object under study. If individual lightweight and stabilized telescopes of 5 m or more in diameter could be moved in this way, then the array would also be a good general purpose telescope. A fundamental question is the image quality that can be achieved in telescopes close to flat ground with no domes.

Making mirror elements

A central issue of any plan for a very large telescope must be a technique to fabricate the huge reflecting surface at acceptable cost and in a reasonable period of time. Without new ideas, mirror costs could take up most of the entire budget. If very low expansion materials are to be used then any structures other than solid sheets are probably ruled out because of very expensive fabrication costs. Costs for material alone are in the range \$50–100 kg⁻¹.

Borosilicate glass honeycomb is an attractive alternative for large mirror elements. The raw material is inexpensive and can be easily fabricated into complex structures by casting at relatively low temperatures. Figure distortion is a problem in solid borosilicate glass mirrors because temperature gradients persist for a long time, but will not be significant in a structure that can rapidly be made isothermal. The Palomar 5-m, the Lick 3-m and the KPNO 2.1-m mirrors are ribbed Pyrex structures with thickness of faceplate and ribs of ~10 cm. Thermal distortion of these mirrors is generally below the 1 arc s level³¹. At Steward Observatory we have undertaken to make large blanks consist-

ing of 3-cm faceplates and still thinner wall honeycomb sandwiched between. These will be ventilated internally with air at ambient temperature, so equilibration will be rapid. A fundamental advantage is that mirror seeing, discussed earlier, will be eliminated. Two 60-cm test mirrors have been made from borosilicate glass: one has a slotted strut type of construction³² and has been figured to a $\lambda/8$ r.m.s. sphere. The other, a similar honeycomb sandwich but made by a single casting in a complex mould, is shown in Fig. 3. Thermal relaxation and structural properties of both mirrors are being investigated by interferometry. Efforts are now directed towards casting in one piece

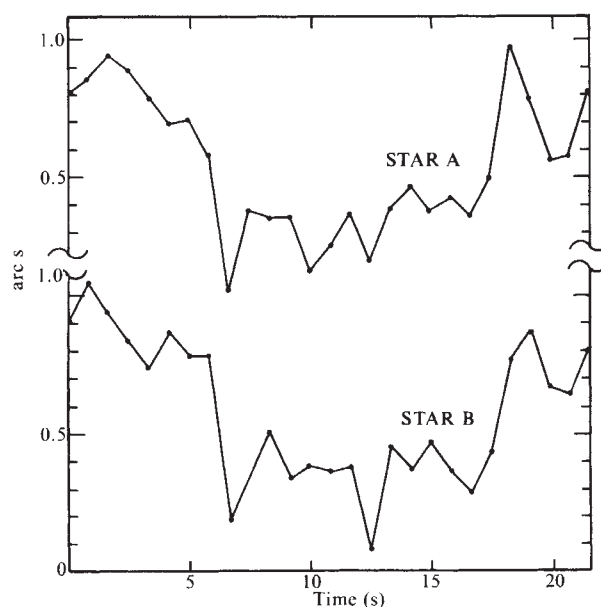


Fig. 4 Atmospheric-induced motion in one coordinate of two stars close to the North Pole separated by 50 arc s. Each data point gives the centre of gravity of an image integrated for 1/6 s. The r.m.s. fluctuation of the difference in positions is 0.11 arc s (ref. 41).

a honeycomb-faceplate sandwich mirror 1.8 m in diameter that will be figured as a parabola and tested alongside the fused silica mirrors in the MMT. We hope to work up to a 7.5-m blank over the next few years.

Ideas for reducing cost are needed in optical figuring. Traditional methods of grinding and polishing are time consuming and expensive for aspherics. One solution was devised by Schmidt, to make the aspheric corrector plates for his camera. A glass blank is polished spherical by conventional methods while being stressed in such a way that on release it takes up the required figure. Lemaitre³³ has demonstrated this method for making on-axis paraboids, making a diffraction limited $f/2$ parabola of 40 cm diameter. Nelson *et al.*²² have extended the technique to make a precise off-axis parabola 35 cm in diameter, stressing the blank during manufacture with levers at the edge. The major effort of Kitt Peak National Observatory in the field of large telescopes is directed towards making matched off-axis elements 2m in diameter by the same method³⁴. Precise tolerance in curvature (~ 1 in 10^5) is needed if a large parabola is to be made from separately figured segments. Another area of concern³⁵ is how these elements are to be made hexagonal, so they can be fitted together to make a nearly seamless surface of low thermal emission. Trimming a large thin circular mirror could disturb its figure as strain is relieved.

Traditional methods of figuring glass are based on the fact that two rigid blocks rubbed against each other with abrasive between will develop surfaces that are very accurately spherical. Machines to do the rubbing need not have any great precision, they only have to drag one piece across the other in a suitable pattern. A different approach is used in the production of optical surfaces by diamond turning of metal substrates. This is done with an extremely accurate lathe, and a single diamond tool of large radius advanced gradually so as to leave a mirror-like finish with only very shallow grooves. Optical elements up to a metre or so with surface accuracy of the order of micrometres have been produced by this method. Polishing can then give surfaces with better than a wave accuracy and remove the small-scale structure³⁶.

Unfortunately, metal substrates have not yet been demonstrated that can withstand the repeated temperature cycling of mountain top observatories without gradually losing their shape. Glass, while stable, cannot be finished like metals with a single diamond, but precision direct machining is possible with an abrasive diamond grinding wheel. This can again bring the surface to within a micrometre or so of the final figure, minimizing the work to be done in polishing. Large mirrors or segments would be turned on an air bearing about a vertical axis, and the tool moved across the surface as in Leighton's³⁷ method of making precision radio telescopes. Parks and Angel³⁸ have worked out a scheme to make mirror segments in which the grinding tool moves on an arm turned about an axis inclined to the vertical, yielding naturally a very close approximation to a parabola.

Electronic image stabilization

Despite the electronic sophistication available, large telescopes are being built which rely exclusively on massive rigid optics and supports to maintain image quality: only the MMT has used active control to replace force rigidity. The potential for achieving very large collecting areas by this technique at relatively low cost is enormous: the MMT has been built for about one-quarter the cost of its rigid equivalent.

There are various ways that active correction might be made. For single-mirror telescopes (or individual elements of a bigger telescope) we can alter the forces that support the primary, or leave it alone and devise a mechanism to distort the secondary or subsequent mirrors. For high frequency control, Hardy's scheme³ uses piezo-electric actuators acting on the back of a thin mirror. Merkle *et al.* have demonstrated³⁹ an electrostatic force system. Magnetic actuators could be used for correction on a less rapid time scale.

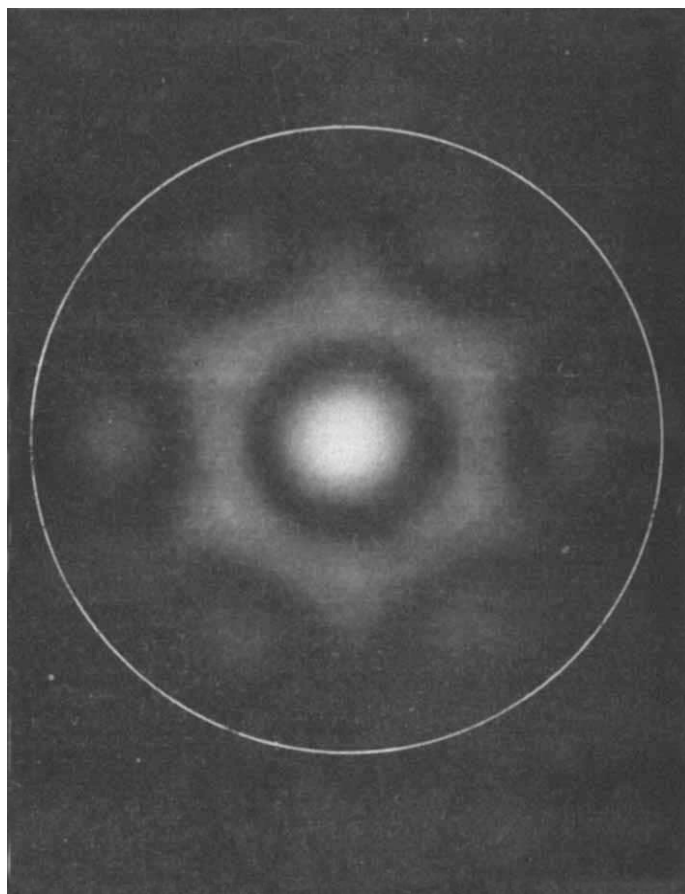


Fig. 5 The diffraction pattern computed by Meinel *et al.* for a phased circular array of six mirrors separated by 0.15% of their diameter. 40% of the total energy is within the first dark ring. The peaks in the continuous hexagonal ring reach 0.11 the intensity of the central peak, the six isolated peaks 0.06. The Airy pattern of one individual mirror, with a first dark ring of diameter shown by the white circle, has a resolution nearly four times poorer than that of the phased pattern.

To achieve a single focus in telescopes that have more than one collecting mirror we need to sense and correct the alignment of each mirror. In the MMT the tilt of individual secondary mirrors is controlled by micrometers driven by stepper motors, giving single steps of 0.05 arc s at the combined focus. Gabor⁴⁰ has demonstrated a ball screw actuator suitable for tilting 2-m primary elements with even higher angular precision.

How do we derive the information necessary to make corrections? Elastic deflections arising from differing orientation with respect to the gravitational field are repeatable, and tables of corrections necessary for their compensation can be stored in the control computer. There will remain unpredictable errors, such as wind forces, uncompensated thermal gradients in the structure, and friction in mirror supports, and in the MMT these residual effects amount to about 1 arc s in image motion⁹ and a few tens of micrometres in pathlength error as the telescope is moved to different parts of the sky³⁰. The residuals which might be achieved in a 15-m class instrument, probably not much worse than in the MMT, are comparable with the random wavefront error already present from atmospheric turbulence.

To realize fully the potential of active optics, one must determine how rays from a star illuminating different parts of the reflecting surface come together at the focus, and feed back corrections to the active elements. This will correct both the structural errors and the large-scale wavefront errors, if done with a frequency response of a few hertz. One method is by

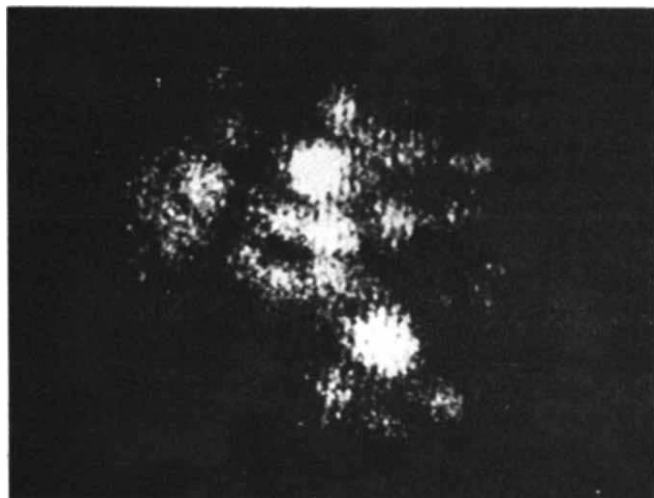


Fig. 6 Interference fringes spaced by 30 m arcs in the highly magnified image of an unresolved star. The image was formed by two of the 1.8-m mirrors of the MMT, with a centre-to-centre spacing of 5 m. (Courtesy of E. K. Hege and P. A. Strittmatter.)

automating the knife edge test: in its basic form the star image is focused on a sharp edge, such as a razor blade, and an auxiliary lens is used behind the edge to form an image of the primary mirror. A tilt of the wavefront in a particular region will then show as a bright or dark patch in the image. Hardy⁴¹ has used a scheme in which an oscillating grating cuts across the star image, and the pupil is imaged on a bank of photomultipliers. Phases of the signals from each photomultiplier then give the tilt error of each part of the pupil. A variant of this idea is where the reference star is imaged on a sharp four-sided pyramid, and each reflected beam passes through a Fabry lens to image the pupil on one of the four sensitive imaging detectors⁴. A complete map across the pupil of tilt of the wavefront in both directions is obtained by intercomparison of the four pupil images. An alternative method that does not use a knife edge uses a lens placed after the combined focus to image the pupil on an array of tilted mirrors. Each mirror directs light from a part of the pupil to a field reimaging lens, and imaging detectors monitor the full field image from each pupil segment. Image locations relative to a reference dot in the combined focal plane then give the tilt of each pupil segment. In both these schemes the problem of maintaining co-alignment of a multiple element telescope reduces to the same problem addressed by offset guiders in today's single element telescopes. An automated system has to be reproduced for each section of the primary that will be independently corrected.

For the wavefront error correction to work, one must either use part of the light from the program star itself, or use a field star close enough for it to share the same large-scale wavefront errors. The angular separation α of shared motion is approximately d/h , where d is the scale of the wavefront error and h is the effective height of the wavefront disturbance. For $d = 2$ m and $h = 4$ km (ref. 42), α is about 2 arc min. Well correlated motion of two stars 50 arc s apart observed with a 1.8 m aperture in conditions of reasonably good seeing (1.2 arc s) is shown in Fig. 4 (ref. 43).

Because much of the work of a large telescope will be with faint objects giving no useful signal for electronic stabilization, the capability for continuous fine tuning of the image relies on finding a guide star within the field of common motion. Fortunately, the development of CCD detectors of excellent red sensitivity means that adequate signals can be obtained. Consider, for example, stars of R magnitude <16 , which with a red filter would stand out clearly even in full moonlight. At the sparsest regions near the galactic poles these have a density of

382 per square degree⁴², and will thus be found with 75% probability within a 2 arc min radius. The light from such a star collected by a 2-m element will give a signal of 8000 electrons s^{-1} in a CCD, adequate to implement the schemes outlined above with a frequency response of a few hertz.

When rapid wavefront tilt corrections are made to a telescope with continuous primary and secondary surfaces then instantaneous large-scale phase errors are also removed. As discussed earlier, this should allow diffraction-limited performance to be achieved in the IR. Even if the full collecting aperture is made from separate elements it should still be possible to maintain correct phasing from tilt measurement provided that the elements are close to each other, and that no steps are allowed to develop in the wavefront going from element to element. Relative phases between elements can be established by occasional observations of broad band interference in an unresolved reference source³⁰. For a system of adjacent segments making a single surface, capacitive sensors such as those developed by Gabor⁴⁰ can be used to maintain accurately steps between adjacent mirrors.

The diffraction pattern of MMT type telescopes when multiple circular apertures are combined perfectly in phase has been explored elsewhere^{25,45,46}. The intensity at the central maximum of the pattern depends only on the total collecting area; at this point all the radiation is added in phase no matter where it was collected. The width of the central peak depends on the diameter of the full array. An exact calculation of a typical diffraction pattern is shown in Fig. 5. This is for a circular array of six mirrors, separated by 0.15% of their diameter. Such a close array could potentially be maintained in phase from tilt sensing. The power of such a telescope for IR imaging is realized when we consider the potential resolution at 10 μm . Equation (1) shows that the phase errors over individual 6-m mirrors will get to $\lambda/10$ or less in good seeing. An MMT configuration of six such mirrors with active correction of phase differences between mirrors would give images close to the diffraction limit of Fig. 5, that is 0.11 arc s FWHM of the central peak, the same as for a filled aperture of 22-m diameter.

Although active wavefront correction to achieve directly diffraction limited images for visible light is not generally possible, various interferometric techniques can be used to study spatial structures at the diffraction limit. Very large single mount telescopes will allow resolution of ~ 5 m arcs at 0.5 μm . An example of interferometry between the full surface of two of the 1.8-m mirrors of the MMT is shown in Fig. 6. The centres of the mirrors are separated by 5 m, giving a fringe spacing of 30 m arcs at the observed wavelength of 0.75 μm . A specialized interferometric telescope array, such as Labeyrie is building, will give longer baseline and higher resolution than any single large telescope, but the latter will be able to reach considerably fainter objects⁴⁷.

Instrumentation and operation

We have focused on current ideas for making the large telescopes which, nevertheless, have excellent images; it is beyond our scope to review other characteristics, but we can indicate some of the challenges.

The sensitivity of present large telescopes in the wavelength range below 1 μm is being brought close to the theoretical maximum using silicon arrays. CCDs have near unit quantum efficiency and negligible noise for nearly all applications. We must also plan to use these detectors in very large telescopes: the challenge is to handle the inevitably large detector area. Because optics of present telescopes and spectrographs are close to the fastest practical speeds, increases in detector area proportional to the primary collecting area are inevitable. Detector areas from 10 to 1,000 cm^2 will be needed, depending on the number of spatial and/or spectral resolution elements to be resolved in a single exposure. Just as we will probably make large collecting areas from more than one piece of glass, we will probably need to make the detector from more than one piece of

silicon. Many options exist for arranging detector chips in close mosaics, or in multiple smaller instruments. Some of the factors are discussed in ref. 18.

Another challenge is to maintain a sizeable field of view for a very large telescope. With good detectors there will be an important role for precise photometry and objective spectroscopy⁴⁸ of many thousands of objects at once in a wide field. With fibres automatically positioned in the focal plane and feeding the light along the slit of a spectrograph, dozens of objects can be observed over a wide field at once, with high spectral resolution if required^{49,50}. If the primary focal ratio is made very fast to keep down tube length and dome size in a single-surface telescope, difficulties can arise in obtaining a wide field. However, two-element reflecting correctors offer a remarkable opportunity of obtaining a large corrected field with a fast primary. An optimized design²³ of the Baker–Paul type^{51,52} explored in detail by ray tracing has an $f/1$ primary and an $f/2$ corrected focus. The achromatic images are no worse than 0.2 arc s over a 1° field.

The third challenge in a giant telescope is to maximize the operational efficiency. The best use is to match the type of observations to the conditions prevailing. Moonlight and twilight are obvious factors but there are also unpredictable changes in transparency, seeing, water vapour, wind and clouds. Ideally one needs several complementary instruments, ready for use with only a few minutes time lost for interchanging. Astronomers need not be present at the site for observations with standard equipment: they could be called to a local data terminal when their observations were to be made. Experiments in this type of operation have already begun at Kitt Peak National Observatory³.

Multiple detectors, automated multiple object probes, automatic interchange of instruments, resident observing staff and communication with remotely located astronomers are expensive. When we consider, though, the capital investment in a 15-m class telescope, the costs of development, construction and operation of these instruments will not be disproportionate.

Conclusion

We are seeing the beginning of a new generation of large telescopes for optical astronomy, characterized by innovative methods of mirror fabrication and active feedback to maintain

image quality. Large, rigid mirror substrates of honeycombed glass are being developed, along with techniques for inexpensive manufacture of precision aspheric surfaces. Advances in sensors and data processing now make it practical to measure continuously atmospheric distortion and make rapid corrections to the wavefront to improve image quality. Very large telescopes incorporating these mirrors and active control should work better and cost less than if built by conventional means. Remarkable advances over current levels of performance, beyond simply collecting more photons, now seem within reach. A 15-m telescope can potentially, with a modest degree of wavefront correction, give images at $10\ \mu\text{m}$ approaching 0.1 arc s in diameter. This is comparable to the diffraction-limited resolution of the space telescope in visible light. Interferometry with the 15-m telescope at visible wavelengths will allow the exploration of structures in faint objects at 5 m arc s resolution, approaching the resolution achieved at radio wavelengths with interferometry over intercontinental baselines.

These unique advantages for higher spatial resolution argue that the first of the very large instruments should be a phasable design. However, when several very large telescopes are to be built it may be advantageous to design different complementary specialized types. For example, a very large telescope directed towards spectroscopy at shorter wavelengths might take the form of an array of independent smaller telescopes, or of a fixed spherical dish with tracking foci, and could be conveniently located on a lower mountain. Telescopes for thermal IR and submillimetre work need to be as high as possible for low water vapour and put a premium on the largest possible phased aperture. Again, for high resolution interferometry widely spaced large telescope elements are needed that move smoothly along the ground. Each type has a place.

This review has benefited from discussions with colleagues at the University of Arizona and the Smithsonian Institution. In particular, Nick Woolf inspired the discussion of atmospheric effects and has made many suggestions for improving the manuscript. I also thank Jacques Beckers and his staff at the MMT, for it is at this telescope that we are working out many of the basic new ideas common to proposed advanced telescopes. Work was supported by the NSF under grant AST 78-22714 and NASA NAGW-121.

- Townes, C. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 797 (Kitt Peak National Observatory, Tucson, 1980).
- Born, M. & Wolf, E. *Principles of Optics* 5th edn (Pergamon, Oxford, 1975).
- Hardy, J. W. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 535 (Kitt Peak National Observatory, Tucson, 1980).
- Woolf, N. J. & Angel, J. R. P. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 1062 (Kitt Peak National Observatory, Tucson, 1980).
- Walker, M. F. *P.A.S.P.* **83**, 401 (1971); *Trans. IAU XVI A*, Pt 1, 219 (1972).
- Barletti, R., Cepatelli, G., Paterno, L., Righini, A. & Speroni, N. *J. opt. Soc. Am.* **66**, 1380 (1976).
- Woolf, N. J. *A. Rev. Astr. Astrophys.* **20** (in the press).
- Carleton, N. P. in *The MMT and the Future of Ground Based Astronomy*, 37 (ed. Weekes, T. C.) (SAO Spec. Rep. No. 385, 1979).
- Beckers, J. *MMT tech. Rep.* No. 10, 10 (1981).
- Lowne, C. M. *Mon. Not. R. astr. Soc.* **188**, 249 (1979).
- Hege, E. K., Angel, J. R. P., Weymann, R. J. & Hubbard, E. N. *Nature* **287**, 416 (1980).
- Lewis, W. C. *Workshop on Optical Fabrication and Testing*, Tucson, 5 (Optical Society of America, 1979).
- Epps, H., Rank, D. & Wampler, J. *Report to the Director of Lick Observatory on the Choice of Primary Mirror for the UC Ten Meter Telescope* (1979).
- Mack, B. *The Observatory* **101**, 135 (1981).
- Pearson, E. *Internal Rep.* Univ. Texas (1981).
- Angel, J. R. P. & Hill, J. *Proc. ESO Conf.* (1981).
- Angel, J. R. P., Arganbright, D., Harmonson, L., Hill, J. M. & Woolf, N. *Proc. IAU Colloq.* No. 67 (1981).
- Disney, M. *Optical and Infrared Telescope for the 1990's* (ed. Hewitt, A.) 649 (Kitt Peak National Observatory, Tucson, 1980).
- Gillette, F. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 293 (Kitt Peak National Observatory, Tucson, 1980).
- Learner, R. C. M. *The Observatory* **101**, 137 (1981).
- Nelson, J. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 11 (Kitt Peak National Observatory, Tucson, 1980).
- Nelson, J., Gabor, G., Hunt, L., Lubliner, J. & Mast, T. *Appl. Opt.* **19**, 2341 (1980).
- Angel, J. R. P., Woolf, N. J. & Epps H. *Pap. S. P. I. E. Meet. Advanced Technology Optical Telescopes* Tucson (1982).
- Beckers, J. *et al. Telescopes of the 1980's* (Annual Reviews Monograph, 1981).
- Meinel, A. *IAU Symp.* No. 27, 611 (1965).
- Drake, F. *Optical and Infrared Telescope for the 1990's*, (ed. Hewitt, A.) 649 (Kitt Peak National Observatory, Tucson, 1980).
- Mertz, L. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 957 (Kitt Peak National Observatory, Tucson, 1980).
- Aikens, R. *et al. NGT Rep.* No. 2, (Kitt Peak National Observatory, Tucson, 1977).
- Angel, J. R. P. *Proc. IAU Colloq.* No. 67, (1981).
- Low, F. J. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 825 (Kitt Peak National Observatory, Tucson, 1980).
- Bowen, I. *IAU Symp.* No. 27, 1313 (1965).
- Loytly, E. Y. & DeVoe, C. F. *I.E.E.E. Trans. Aerospace Electron. Syst.* **5**, 300 (1969).
- Lemaître, G. *Optical and Infrared Telescopes for the 1990's* (ed. Hewitt, A.) 896 (Kitt Peak National Observatory, Tucson, 1980).
- Ghoseil, I. *Optical and Infrared Telescope for the 1990's* (ed. Hewitt, A.) 616 (Kitt Peak National Observatory, Tucson, 1980).
- Brown, D. *Meet. New Techniques and Telescopes in Optical Astronomy* (Royal Astronomical Society, London, 1980).
- Parks, R. E. *Workshop on Optical Fabrication and Testing*, Tucson (Optical Society of America, 1979).
- Parks, R. and Angel, J. R. P. *Paper S.P.I.E. Meet. Instrumentation in Astronomy-IV*, Tucson (1982).
- Leighton, R. B. *Final Techn. Rep. NSF Proj. AST 73-04908; Astrophys. J.* **230**, 149 (1978).
- Merkle, F., Freischlad, K. & Bille, J. in *Scientific Importance of High Angular Resolution at Infrared and Optical Wavelengths*, Munich (1981).
- Gabor, G. *S.P.I.E. Proc. Instrumentation in Astronomy III* **172**, (1979).
- Hardy, J. W. *Proc. ESO Conf. on Optical Telescopes of the Future*, Geneva, 455 (1977).
- Young, A. T. *Astrophys. J.* **189**, 587 (1974).
- Angel, J. R. P. *Proc. ESO Conf. Scientific Importance of High Angular Resolution at Infrared and Optical Wavelengths*, Munich (1981).
- Bahcall, J. & Soneira, R. M. Preprint (Princeton Univ., 1981).
- Meinel, A. B., Meinel, M. & Woolf, N. J. *J. Appl. Opt. opt. Engag* **9** (in the press).
- Sanger, G. thesis, Univ. Arizona (1978).
- Roddiar, F. in *Scientific Importance of High Angular Resolution at Infrared Optical Wavelengths* (eds Ulrich, & Kjar.) (ESO, 1981).
- Hoag, A. A. & Schroeder, D. J. *P. A. S. P.* **82**, 1141 (1970).
- Hill, J. M., Angel, J. R. P., Scott, J. S., Lindley, D. & Hintzen, P. *Astrophys. J. Lett.* **242** L69 (1980).
- Angel, J. R. P. in *The MMT and the Future of Ground-Based Astronomy* (ed. Weekes, T. C.) (SAO Special Report No. 385, 1979).
- Paul, M. *Rev. Opt.* **14**, 169 (1935).
- Baker, J. G. *I.E.E.E. Trans. Aerospace Electron. Syst.* **5**, 261 (1969).
- Janes, C., Robinson, W. & Shechter, P. *KPNO tech. Rep.* (in preparation.).
- Hege, E. K., Hubbard, E. N., Strittmatter, P. A. & Worden, S. P. *Astrophys. J. Lett.* **248**, L1 (1981).
- Woolf, N. J., Angel, J. R. P., Antebi, J., Carleton, N. & Barr, L. D. *Paper S. P. I. E. Meet. Advanced Technology of Optical Telescopes*, Tucson (1982).

ARTICLES

Simulating silicate structures and the structural chemistry of pyroxenoids

C. R. A. Catlow*, J. M. Thomas†, S. C. Parker* & D. A. Jefferson†

* Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

† Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK

Chain silicates, both naturally-occurring and synthetic, exhibit 'backbone repeat' patterns of corner-linked SiO_4 tetrahedra. The extent to which these repeats may be interpreted or predicted, as a function of the associated cation M^{2+} (in the general formula MSiO_3), is discussed, and the broader issues involved in the theoretical simulation of the structures of inorganic silicates are considered.

NATURALLY-occurring and synthetic silicates display an extensive range of structural types including isolated anions, chains, rings, sheets and various kinds of framework, all based on SiO_4^{4-} tetrahedra¹⁻³. The diversity of structural types is exceptional, being matched only by the borates, where both trigonally and tetrahedrally bonded atoms may occur⁴. Even so, the total number of known silicate structures is substantially less than that theoretically predicted on simple topological arguments. Several attempts^{5,6} have been made recently to rationalize and interrelate the structures of the silicates using largely qualitative solid-state, chemical principles. We have pursued a quantitative approach with the hope, ultimately, of predicting the behaviour of many different kinds of novel silicates. Such computations also offer an attractive method of assimilating new structural information brought to light⁷⁻¹¹ by real-space crystallography, using high-resolution electron microscopy (HREM). In particular, we have sought to interpret the structure composition relationships that have emerged from HREM studies of the so-called pyroxenoid silicates^{12,13}.

The pyroxenoids possess a general formula MSiO_3 ($M \equiv \text{Ca}^{2+}$, Fe^{2+} , Mn^{2+} , Mg^{2+} and so on). They display intriguing variability in their structural types and inter-relationships. The difference between individual members of this family manifests itself in the backbone repeat patterns of the single chain of corner-sharing SiO_4^{4-} tetrahedra.

Nature of the experimental facts

Figure 1 shows the various kinds of backbone repeat patterns characteristic of the diopside ($\text{CaMg}(\text{SiO}_3)_2$), wollastonite (CaSiO_3), rhodonite ($(\text{Mn}, \text{Ca})\text{SiO}_3$) and pyroxmangite ($(\text{Mn}, \text{Fe}, \text{Ca}, \text{Mg})\text{SiO}_3$) structures. In wollastonite, three SiO_4^{4-} tetrahedra constitute the unit (7.3 Å in length); in rhodonite, the five-unit repeat runs to 12.5 Å; whereas in pyroxmangite the seven-unit repeat is 17.4 Å. Diopside, with a two-unit repeat (5.2 Å), is an archetypal member of the pyroxene family of chain silicates: the other three are pyroxenoids.

X-ray studies of naturally-occurring and synthetic analogues of pyroxenoids are fraught with difficulties principally because of the occurrence of complicated intergrowths of various members of the pyroxenoid range of structures. Great headway can, however, be made¹⁴⁻¹⁷ by real-space direct methods (HREM) as may be seen in Fig. 2. Although it is difficult to be sure that equilibrium states have been reached, certain preferences for given pyroxenoid structures can be noted for given cations (or stoichiometric ratio of cations). Thus, as the synthetic pyroxenoid of formula $\text{Mn}_3\text{MgSi}_4\text{O}_{12}$, emerges from the glassy state, the pyroxene structural type is first seen to predominate, with embryonic regions (a few unit cells in width) of various

other pyroxenoids (Fig. 2a). Eventually (after prolonged annealing at 800 °C) a marked propensity for the pyroxmangite structural type prevails (Fig. 2c); but an intermediate stage (Fig. 2b) may be detected by HREM in which strips of wollastonite are inserted regularly to produce small regions of pyroxmangite.

Likewise the synthetic pyroxenoid of formula $\text{MnFeSi}_2\text{O}_6$ first displays some individual strips of nine-unit backbone repeats (that is the ferrosilite III structure which has a unit length of 22.7 Å) interspersed amongst pyroxmangite and rhodonite structural types (Fig. 3a). Gradually (Fig. 3b), five-unit repeats gain ascendancy, and a well-ordered rhodonite (five-unit repeat) ultimately prevails.

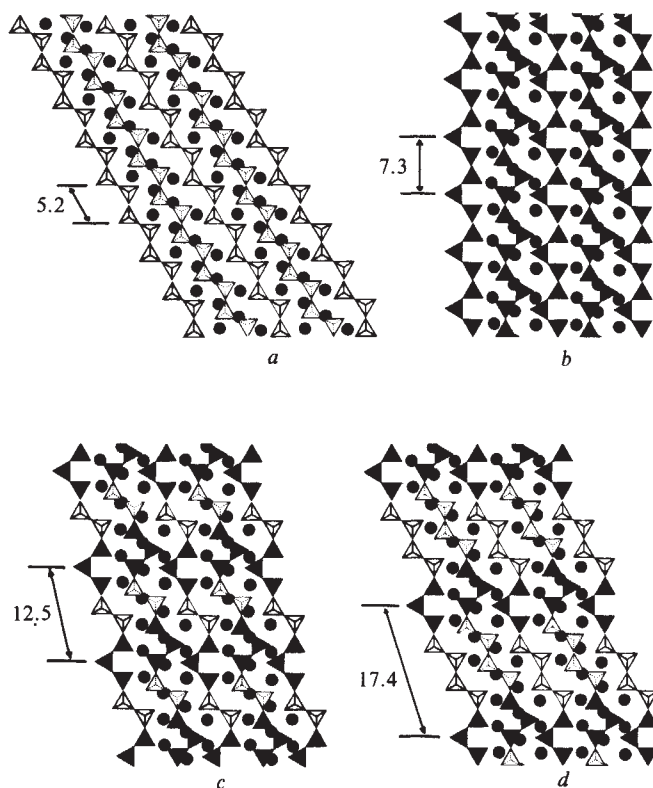


Fig. 1 Illustration of the pyroxenoid structures referred to in the text. a, Diopside (two unit repeat); b, wollastonite (three unit); c, rhodonite (five unit); d, pyroxmangite (seven unit).

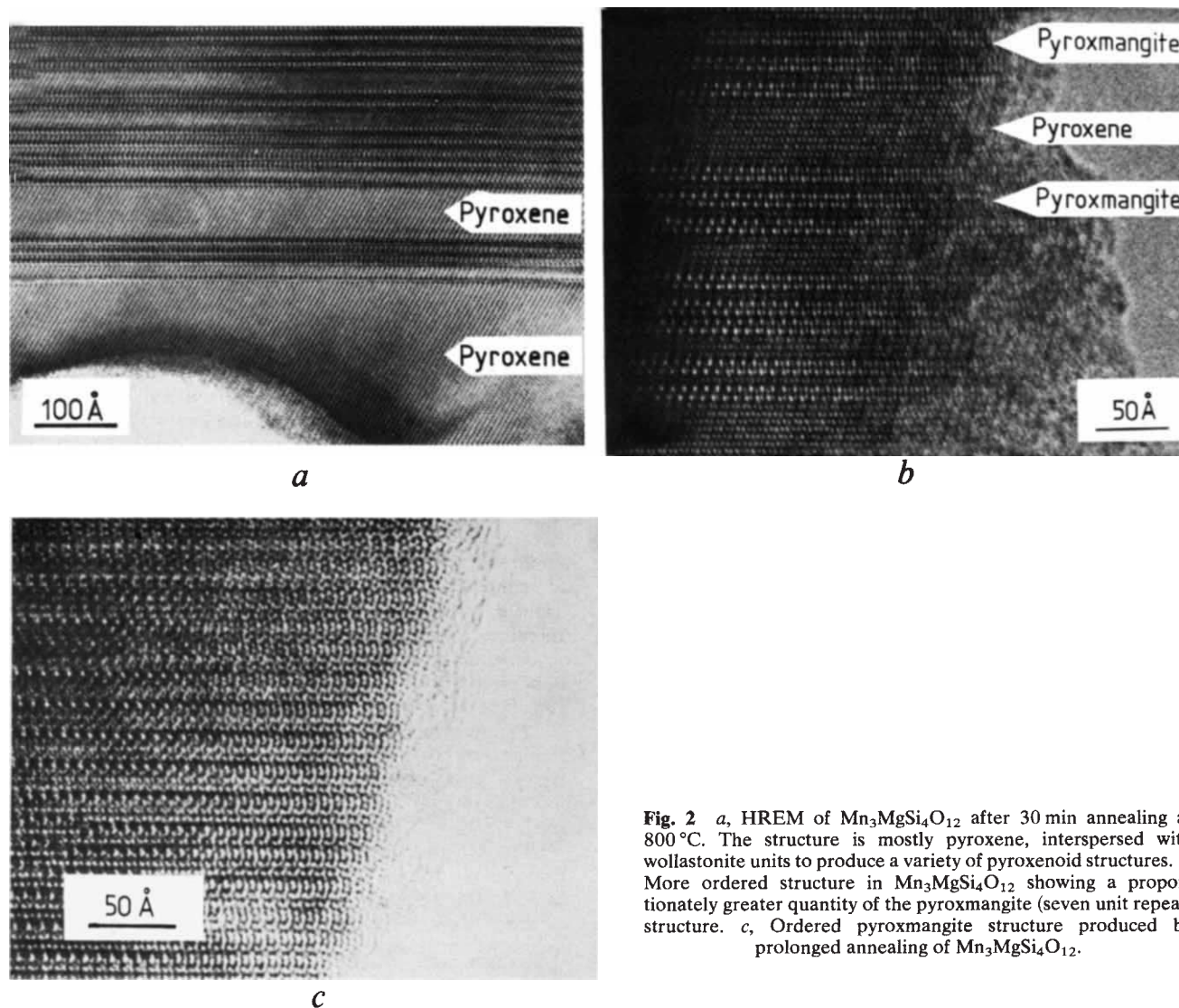


Fig. 2 *a*, HREM of $\text{Mn}_3\text{MgSi}_4\text{O}_{12}$ after 30 min annealing at 800 °C. The structure is mostly pyroxene, interspersed with wollastonite units to produce a variety of pyroxenoid structures. *b*, More ordered structure in $\text{Mn}_3\text{MgSi}_4\text{O}_{12}$ showing a proportionately greater quantity of the pyroxmangite (seven unit repeat) structure. *c*, Ordered pyroxmangite structure produced by prolonged annealing of $\text{Mn}_3\text{MgSi}_4\text{O}_{12}$.

One key question to which our simulation studies are addressed is which structural type is preferred for a particular cationic sublattice? A related question raises the issue of whether mixed cations (ordered or disordered?) can tilt preference from one structural type to another?

Knowing the success that procedures using pair potentials have achieved in interpreting the behaviour of molecular crystals, and organic solid-state chemistry^{18–21}, it seemed appropriate to evolve computationally more advanced procedures that could cope with the demanding set of circumstances associated with the silicates. Fortunately, the techniques developed by C.R.A.C. and Norgett^{22–25}, which have proved successful in the study of non-stoichiometry, extended defects and superionic conductivity, have considerable value in the present context and promise much more extensive interpretive scope in silicate chemistry. It has been shown^{22–32} that simulation techniques can elucidate the nature of the complex defect aggregation modes in heavily defective systems such as Fe_{1-x}O (ref. 26) and UO_{2+x} (ref. 27), and identify the factors that determine whether compositional variation is accommodated by extended rather than by point defects^{28,29}.

The simulation procedure

The methods are, at least conceptually, simple. A potential model is used to describe the interactions between the ions in the crystal; the energy of the crystal is then minimized with respect to all structural parameters. [The methods do not explicitly include the effects of thermal motion. However, this omission is

unlikely to affect seriously the use of the methods for structural predictions.] In contrast to this conceptual simplicity the actual computational techniques are sophisticated. This relates first to the handling of the summation procedures, where we have developed procedures available in the PLUTO lattice simulation code²⁶ to evaluate the long-range electrostatic interactions. (Our methods are based on the Ewald transformation which replaces a slowly convergent summation in real space with a rapidly convergent reciprocal space sum). Second, efficient minimization methods are vital if we are to cope with large unit cells. Earlier work showed the advantages of Newton–Raphson procedures which result in rapid and stable convergence³⁰. Special problems arise when the minimization procedure adjusts unit cell dimensions in addition to ion coordinates—an essential feature in the simulation of silicates.

Application of the techniques to silicates requires first, development of adequate potential models for these materials and second, adaptation of minimization procedures to handle the complex structural problems posed by these systems.

Potential models for the silicates

Successful simulation studies of transition metal and actinide oxides have been based on ionic model potentials with two-body, short range potentials described by a simple analytic function (generally of the Born–Mayer or Buckingham form). Although the ionic model is clearly an approximate description of the bonding in these oxides, there is growing evidence (which includes the success of the calculations) that the approximation

Table 1 Short-range parameters

$V(r) = A \exp(-r/\rho) - C/r^6$			
<i>a</i>	<i>A</i> (eV)	ρ (Å)	<i>C</i> (eV Å ⁶)
O ²⁻ –O ²⁻ *	2,2764.3	0.1490	27.88
Mg ²⁺ –O ²⁻ †	2,214.39	0.2756	4.45
Ca ²⁺ –O ²⁻ †	1,996.35	0.3189	26.57
Si ²⁺ –O ²⁻ †	2,187.61	0.3377	64.08
Mn ²⁺ –O ²⁻ †	2,618.35	0.3033	27.46
Fe ²⁺ –O ²⁻ †	2,886.3	0.2973	29.73
<i>b</i>			
Si ⁴⁺ –O ²⁻	998.98	0.3455	0.0

* See refs 27, 39.

† See ref. 44.

is good for materials such as TiO₂ (ref. 31) NiO (ref. 32) and UO₂ (ref. 27). To what extent, however, has the model any validity when applied to quartz and to silicates? To answer this question we must first consider the meaning and implication of the term ionicity.

Following Cochran³³ we note that there are two distinct (although related) operational definitions of this concept. The first concerns the distribution of valence shell electron density, and although we have no definite evidence, use of this criterion would almost certainly suggest the inapplicability of the ionic model to silicates. The second criterion, which has much greater relevance to the present study, is in terms of interatomic forces and the response of the crystal to external perturbations (elastic or electrostatic) or internal structural modifications (such as the formation of defects or the introduction of impurity ions). In this context a solid is regarded as ionic if a description of the lattice based on ions can correctly predict (1) the lattice energetics of the solid, (2) its elastic and dielectric properties and (3) the formation energies of defects and structural response of the lattice to the introduction of disorder. This latter criterion is equally valid to, but distinct from, the former. It is essentially a definition in terms of forces rather than charge distribution. In the context of the debate on the extent of ionicity in solids, it is important that this distinction in the connotation of the term should be appreciated. We shall show below that the ionic model forms a reasonable basis for the simulation of silicates.

We have therefore developed ionic model potentials which are generally applicable to ring and chain silicates. Our procedure differs from that adopted in earlier studies of simpler inorganic oxides^{27,29,32} where variable parameters in a general potential model were adjusted so as to give the best fit to experimental elastic and dielectric data. Such data are in general unavailable for minerals. However, the structures of complex low symmetry crystals are themselves a rich source of interatomic potentials, because, in such crystals a given type of interaction potential (say silicon···oxygen) is generally sampled over a wide range of separations. (This fact has already been successfully exploited³⁴, in developing a potential model for ZrO₂ from the low temperature, low symmetry structure.) We have, therefore, used the structural data available on a range of chain and ring silicates to develop our generalized potential model for these systems.

The precise procedure is as follows. First, we define the form of our potential model which has two main features: (1) the crystal is simulated by point entities of integral charge; and (2) short-range forces are included for the Si···O and O···O interactions (we discovered that inclusion of Si···Si interactions over short-range was not necessary in order to reproduce structural data on ring and chain silicates—a result which is apparently at variance with recent suggestions^{5,35–38} concerning the dominance of these interactions in determining the structures of minerals.) The short-range forces are described by a potential of the form:

$$V(r) = A e^{-r/\rho} - Cr^{-6}$$

For the O···O interaction we took parameters (reported in Table 1a) which were derived from a quantum mechanical calculation of the interactions between two O²⁻ ions^{27,39}. For the interaction between the metal cations and oxygen in the minerals studied, we used parameters obtained for the appropriate metal oxide. All metal–oxygen parameters used are listed in Table 1a. Interactions between metal and Si⁴⁺ ions were ignored, as these are negligible at the interatomic spacings in silicates. The parameters *A*, ρ and *C* for the Si···O interactions were taken as variables and were fitted to the structural data. This procedure involves adjusting the parameters until the forces acting on all atoms in the unit cell are zero, and the unit cell itself is in equilibrium with its observed dimensions. This is achieved by a second application of the PLUTO program, which, given a specified structure and potential, calculates both the forces acting on the individual atoms in the cell and forces of dilation or contraction acting on the unit cell as a whole. The program is linked with a least-squares routine which minimizes these forces with respect to the variable potential parameters. In practice, for relatively simple potentials of the type we have used, it is not possible to derive a fully strain-free potential. Thus, to check that our final fitted potential is in adequate agreement with the structural data, we allow the structure to equilibrate (that is achieve its minimum energy configuration by adjusting all structural parameters) using the potential. Our criterion for the acceptability of the fitted potential was that in the process of equilibration no atom should be displaced from its observed position by more than 0.25 Å. In general the discrepancies were much smaller than this figure (typical values were 0.1 Å). However, the neglect of thermal motions by our calculations when combined with the errors in the crystallographic data could lead to a discrepancy of this magnitude in an extreme case, although in general we would expect a good potential to give better agreement with high quality structural data.

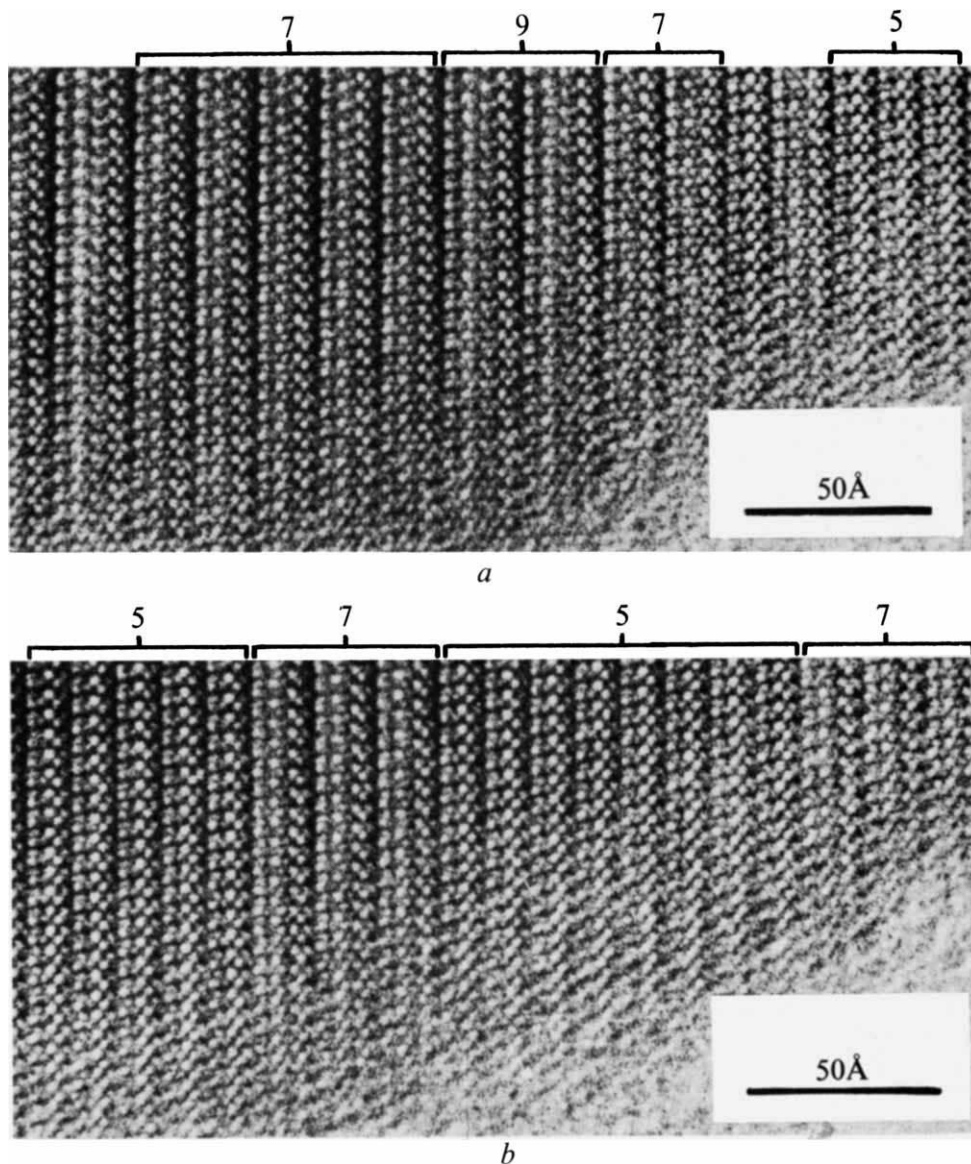
The above procedure was applied to the six silicates listed in Table 2. We found that a single set of parameters listed in Table 1b for the potential could fit the observed structural data for the crystals, which include structures based on isolated SiO₄ units as well as rings and chains. On the basis of the second of the two criteria given below, the result supports the suitability of the ionic model for describing the bonding in these systems. The model yields correct structures for those crystals in which there is a considerable range of structural variation. The implication is that the forces in the silicate lattice are being described correctly, and this conclusion is supported by calculations of the lattice energies. Studies of silicates have been reported using ionic model potentials^{40–43}; and the work of Jenkins *et al.*^{41,42} has clearly demonstrated the accuracy of lattice energy calculations using these potentials. We should note, however, that it is much more difficult to derive potentials which satisfactorily model structural data on more condensed silicate systems, that is layer and framework structures. Investigation of more sophisticated models for these systems is continuing.

We should emphasize the important differences between the approach adopted here and that used in earlier theoretical studies of mineral systems. First our approach is fully quantitative including representations of short range as well as

Table 2 Mineral structures used in calculating short-range parameters for silicon–oxygen interaction

Name	Formula unit	Ref.
Monticellite	CaMgSiO ₄	45
Sillimanite	Al ₂ SiO ₅	46
Sphene	Ca(TiO)SiO ₄	47
Benitoite	BaTi(SiO ₃) ₃	48
Diopside	CaMg(SiO ₃) ₂	49
Jadeite	NaAl(SiO ₃) ₂	49
β-Wollastonite	CaSiO ₃	50

Fig. 3 *a*, Disordered pyroxenoid structure produced by annealing a glass of composition $\text{MnFeSi}_2\text{O}_6$. (The nine-unit repeat has the so-called ferrosilite(III) structure.) *b*, Ordered rhodonite (eight five-unit repeat) structure predominating after annealing of $\text{MnFeSi}_2\text{O}_6$.



electrostatic terms (which are the sole terms considered in several previous studies^{41,43}). Second, structural data are used as a source of interatomic potentials. Third, energy minimization is used in predicting structural properties.

Further information on the viability of our parameterized potential will follow from the study of the structural effects of the variation of the divalent cation present in chain-structured systems. Before presenting the results of such calculations the problem of energy minimization must, however, be considered.

Minimization methods

Two points are of general importance to the simulation of complex crystal structures. The first concerns the adjustment of unit cell dimensions in such calculations. We found that, for example, on varying the nature of the divalent cation present in a pyroxenoid silicate it was inadequate merely to adjust the coordinates of the ions surrounding the divalent cation. It was also necessary to include dilation (or contraction) of the unit cell; and, moreover, the dilation must be allowed to be anisotropic. Indeed, failure to include anisotropic dilation may lead to qualitatively inconsistent results.

The second point concerns the question of the precise minimization techniques. Our general approach was to modify the unit cell dimensions followed by an iterative adjustment of the atomic coordinates until these are in equilibrium with respect to the new cell. Subsequent cycles then effect further variation of

the cell dimension and the procedure continued until a structure is attained which is in equilibrium with respect to both unit cell parameters and atomic coordinates. Note that in adjusting unit cell dimensions it is very important to vary the angle between the axes of the cell as well as the lengths of the cell edges. Failure to include this feature will drastically reduce the efficiency of the minimizations. We should also note that the equilibrated cell

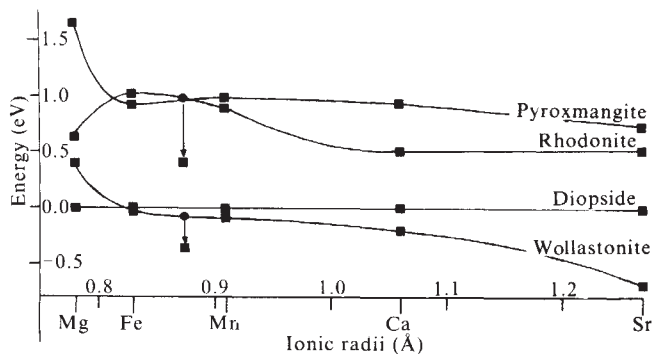


Fig. 4 Plot of energy (per mol SiO_3) relative to that of diopside as a function of radius of the cation M^{2+} . The arrow ($\downarrow$) indicates stabilization arising from the mixed cation effect in Mn_3Mg ; ●, result calculated for average cation radius.

dimensions were within 2% of observed values for all compounds for which crystallographic data are available.

Results and discussion

Calculations were performed on the four pyroxenoid structures, diopside, wollastonite, rhodonite and pyroxmangite. In each case the structures were investigated with the five cations Mg^{2+} , Ca^{2+} , Fe^{2+} , Mn^{2+} , and Sr^{2+} ; a full minimization was performed both with respect to the unit cell dimensions and the ion coordinates. The results are summarized in Table 3, where we give the lattice energy of the equilibrated structures per Si atom; and Fig. 4 displays the calculated energies relative to that obtained for the diopside structure.

The following conclusions may be drawn from these calculations: (1) The pyroxmangite and rhodonite structures are never energetically favoured with respect to the diopside and wollastonite structures. (2) Pyroxmangite and rhodonite have very similar lattice energies especially for Fe^{2+} and Mn^{2+} . (3) The most energetically favoured structure is found to be wollastonite for all ions except Mg^{2+} for which diopside is preferred. The energetic preference for the wollastonite structure increases with the radius of the cation.

The first of the above conclusions does not apparently accord with the experimental data as stable rhodonite and pyroxmangite structures are observed. However, in most cases these are found with mixtures of divalent cations, for example, the structural pyroxenoid $\text{Mn}_3\text{Mg}(\text{SiO}_3)_4$, mentioned earlier, which has the pyroxmangite structure (Fig. 2). We investigated whether these structures could be stabilized by the presence of several, rather than one type of divalent cation. In practice, such calculations could be more conveniently performed on the rhodonite structure with 3/4 Mn^{2+} and 1/4 Mg^{2+} cations. (The Mg^{2+} cations were allocated to the smallest cation sites in the crystal.) The result was compared with that for a crystal in which all the divalent cations are the same with a potential taken as a weighted average of that for $\text{Mn}\cdots\text{O}$ and $\text{Mg}\cdots\text{O}$ interactions. The former was more stable, relative to the diopside structure, by 0.6 eV, than the latter, which gives an estimate of the magnitude of the 'mixed-cation' effect in this structure.

The wollastonite structures were stabilized by only 0.2 eV by this mixed cation effect, when this was calculated in an analogous fashion. Thus, although the relative stabilization of the rhodonite structure by the mixed cation distributed is insufficient in this particular case to render this structure more energetically favourable than wollastonite, the effect is clearly large and may be decisive in certain cases. One situation for which this appears to be so is $\text{CaMg}(\text{SiO}_3)_2$, which when an

Table 3 Calculated lattice energy of MSiO_3 per Si atom

Cation (M^{2+})	Structure			
	Diopside (eV)	Wollastonite (eV)	Rhodonite (eV)	Pyroxmangite (eV)
Mg	163.534	163.133	162.909	161.888
Fe	159.673	159.682	158.582	158.754
Mn	159.204	159.285	158.311	158.232
Ca	158.424	158.622	157.921	157.496
Sr	156.081	156.725	155.564	155.359
Mn_3Mg	159.754	160.137	159.431	—
(Mn_3Mg) composite*	160.216	160.126	159.254	—
CaMg	161.139	160.683	160.130	—
(CaMg) composite*	160.958	160.778	159.926	—

Lattice energies are all calculated as negative quantities; the modulus is given in the table.

* (M_1M_2) indicates a hypothetical ion with anionic radius equal to the mean of M_1 and M_2 .

'average' potential is used, is calculated to be more stable in the wollastonite structure, but when a proper mixed cation distribution is used is found to have greater stability (by 0.6 eV per Si) in the diopside structure as is observed experimentally.

The results for $\text{CaMg}(\text{SiO}_3)_2$ and CaSiO_3 are thus most encouraging. Each is correctly predicted to have the structure that is observed experimentally. Moreover, the trend in the variation with ionic radius of the relative energies of the wollastonite and diopside structures is in accord with experiment. It remains an open question as to whether rhodonite and pyroxmangite structures are ever thermodynamically stable with respect to the wollastonite structure. On the basis of our results we suggest that pyroxenoids are metastable when they crystallize exclusively into either the rhodonite or pyroxmangite structures. Our computations also show that both structures may be stabilized if there is a mixed cation sublattice of the kind discussed above.

We shall show elsewhere that computations of the kind presented here also reveal much valuable information, pertaining to the behaviour of aluminosilicates such as zeolites. In particular the energetics associated with various possible Si, Al, ordering schemes^{51,52} are assessed; and it transpires that the total energy of the aluminosilicates can be a sensitive function of cation position.

We thank the SRC for support and AERE Harwell for generous provision of computer time.

Received 17 July; accepted 29 December 1981.

- Bragg, W. L. & Claringbull, G. F. *Crystal Structures of Minerals* (Bell, London, 1965).
- Belov, N. V. *Crystal Chemistry of Large-Cation Silicates* (Consultants Bureau, New York, 1963).
- Liebau, F. *Handbook of Geochemistry* Vol. 2, Ch. 14 (Springer, Berlin, 1971).
- Waugh, J. L. T. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) 731 (Freeman, San Francisco, 1968).
- O'Keefe, M. & Hyde, B. G. *Acta crystallogr.* **B34**, 27 (1978).
- Dent Glasser, L. S. *Chem. Br.* **18**, 33 (1982).
- Alario Franco, M. A., Jefferson, D. A., Pugh, N. J., Thomas, J. M. & Bishop, A. C. *Mater. Res. Bull.* **15**, 73 (1980).
- Thomas, J. M., Jefferson, D. A., Mallinson, L. G., Smith, D. J. & Crawford, E. S. *Chem. Scr.* **14**, 1677 (1978–79).
- Jefferson, D. A., Thomas, J. M., Alario Franco, M. A. & Pugh, N. J. *Acta crystallogr.* **A36**, 1058 (1980).
- Jefferson, D. A. & Pugh, N. J. *Acta crystallogr.* **A37**, 281 (1981).
- Thomas, J. M. & Jefferson, D. A. *Endeavour, New Ser.* **2**, 127 (1978).
- Thomas, J. M. *New Scientist*, 580 (21 August 1980).
- Shu-Lin Wen, Jefferson, D. A. & Thomas, J. M. *Mater. Res. Bull.* **15**, 1643 (1980); *JCS Chem. Commun.* (in the press).
- Jefferson, D. A. *et al. Nature* **281**, 51 (1979).
- Tilley, R. J. D. *R. Soc. Chem. Spec. Per. Rep.* **8**, 121 (1980).
- Bursill, L. A. *Chem. Scr.* **14**, 83 (1978–79).
- Thomas, J. M., Millward, G. R. & Bursill, L. A. *Phil. Trans. R. Soc. A* **300**, 43 (1981).
- Kitaigorodsky, A. I. *Chem. Soc. Rev.* **7**, 133 (1978).
- Williams, D. E. *J. chem. Phys.* **45**, 3770 (1966).
- Ramdas, S. & Thomas, J. M. *R. Soc. Chem. Spec. Per. Rep. (Chem. Phys. Solids and Surfaces)* **7**, 31 (1978).
- Grammaticioli, C. M. *et al. JCS Faraday Trans II* **76**, 1336 (1980).
- Catlow, C. R. A. & Norgett, M. J. *UKAEA Rep. M2936* (1978).
- Norgett, M. J. *UKAEA Rep. AERE (Harwell)* **R.7650** (1974).

- Catlow, C. R. A. *J. Phys.* **C6–53** (1980).
- Catlow, C. R. A. & Mackrodt, W. C. (eds) in *Lecture Notes in Physics* (Springer, Berlin, 1982).
- Catlow, C. R. A. & Fender, B. E. F. *J. Phys.* **C8**, 3267 (1975).
- Catlow, C. R. A. *Proc. R. Soc. A* **353**, 533 (1977).
- Catlow, C. R. A. & James, R. *Nature* **272**, 602 (1978).
- Catlow, C. R. A. & James, R. *R. Soc. Chem. Spec. Per. Rep. (Chem. Phys. Solids and Surfaces)* **8**, 108 (1980).
- Norgett, M. J. & Fletcher, R. J. *J. Phys.* **C3**, L190 (1970).
- Catlow, C. R. A. & James, R. *Proc. R. Soc.* (submitted).
- Catlow, C. R. A., Mackrodt, W. C., Norgett, M. J. & Stoneham, A. M. *Phil. Mag.* **A40**, 162 (1979).
- Cochran, W. *Nature* **191**, 60 (1961).
- Butler, V., Catlow, C. R. A. & Fender, B. E. F. *Solid State Ionics* **2**, 245–251 (1981).
- O'Keefe, M. & Hyde, B. G. *Acta crystallogr.* **B32**, 2923 (1976).
- Glidewell, C. *Inorg. Nucl. Chem. Lett.* **11**, 353 (1975).
- Glidewell, C. *Inorg. Nucl. Chem. Lett.* **38**, 669 (1976).
- Glidewell, C. *Inorg. Nucl. Chem. Lett.* **13**, 65 (1977).
- Catlow, C. R. A. thesis, Oxford Univ. (also available as UKAEA Rep. TP566 (1974)).
- Baur, W. H. *Am. Miner.* **57**, 709 (1972).
- Jenkins, H. D. B. & Hartman, P. *Phil. Trans. R. Soc. A* **293**, 169 (1979).
- Jenkins, H. D. B. & Pratt, K. F. *JCS Faraday II*, **74**, 968 (1978).
- Alberti, A. & Vezzalini, G. *Z. Krist.* **147**, 167 (1978).
- Mackrodt, W. C. & Stewart, R. F. *J. Phys.* **C12**, 431 (1979).
- Lager, G. A. & Meagher, E. P. *Am. Miner.* **63**, 365 (1978).
- Winter, J. K. & Ghose, S. *Am. Miner.* **64**, 573 (1979).
- Wyckoff, R. W. G. *Crystal Structures* 2nd edn (Interscience, New York, 1968).
- Fischer, K. Z. *Krist.* **129**, 222 (1969).
- Cameron, M., Sueno, S., Prewitt, C. T. & Papike, J. J. *Am. Miner.* **58**, 594 (1973).
- Ito, T., Sadanga, R., Takeuchi, Y. & Tokonami, M. *Proc. Jap. Acad.* **45**, 913 (1969).
- Ramdas, S., Thomas, J. M., Klinowski, J., Fyfe, C. A. & Hartman, J. S. *Nature* **292**, 228 (1981).
- Bursill, L. A., Lodge, E. A. & Thomas, J. M. *Nature* **286**, 111 (1980).

Primary structure of the human Met- and Leu-enkephalin precursor and its mRNA

Michael Comb*, Peter H. Seeburg†, John Adelman†, Lee Eiden‡ & Edward Herbert*

* University of Oregon, Eugene, Oregon 97403, USA

† Genentech Corporation, 460 Point San Bruno Boulevard, San Francisco, California 94080, USA

‡ National Institutes of Health, Bethesda, Maryland 20205, USA

The nucleotide sequence of a complete cDNA copy of enkephalin precursor mRNA from human pheochromocytoma is reported. The corresponding amino acid sequence shows that the precursor is 267 amino acids long and contains six interspersed Met-enkephalin sequences and one Leu-enkephalin sequence. Five of the seven enkephalins are flanked on both sides by pairs of basic amino acid residues. The precursor does not contain the sequences of the opioid peptides, dynorphin, α -neo-endorphin or β -endorphin.

A NUMBER of peptides have been isolated in the past 6 years that exhibit opioid activity. The simplest of these are methionine (Met)- and leucine (Leu)-enkephalin first identified in brain tissue in 1975 (ref. 1)¹. The other opioid peptides are C-terminal extensions of either Leu-enkephalin or Met-enkephalin. For example, β -endorphin is a C-terminal extension of Met-enkephalin^{2,3}, and dynorphin⁴ and α -neo-endorphin⁵ are C-terminal extensions of Leu-enkephalin (Fig. 1).

Because of the structural similarities of these peptides it was postulated that β -endorphin is the precursor of Met-enkephalin and that dynorphin (or α -neo-endorphin) is the precursor of Leu-enkephalin. However, in the past 2 years it has become clear that the enkephalins are derived from a different gene product than β -endorphin, and that β -endorphin is synthesized from a protein, pro-opiomelanocortin (POMC), that contains sequences of adrenocorticotrophic (ACTH) and melanocyte stimulating (MSH) hormones⁶⁻⁸. The source of the enkephalins has been elucidated recently by studying the biosynthesis of these peptides in bovine adrenal medulla^{9,10}. Enkephalin peptides are abundant in adrenal medulla and can be released by neurotransmitters specific for that tissue. Several enkephalin-containing peptides have been purified from chromaffin granules of the bovine adrenal medulla and shown by sequence analysis to contain more than one enkephalin sequence^{9,10}. These studies suggest that a precursor protein exists which contains multiple copies of the enkephalins and that the enkephalins can be released from this protein by digestion with trypsin-like enzymes^{9,10}.

Enkephalins are also abundant in human pheochromocytoma, a tumour derived from the adrenal medulla. The RNA from this tumour contains a high level of enkephalin mRNA sequences as demonstrated by cell-free translation studies¹¹. Therefore, we have used human pheochromocytoma as a source of human proenkephalin mRNA for the preparation and cloning of human enkephalin cDNA.

We have previously shown that a synthetic oligodeoxynucleotide probe complementary to a mRNA sequence that codes for Met-enkephalin can be used to detect a proenkephalin mRNA species in bovine adrenal and human pheochromocytoma by the Northern blot procedure¹². These species of RNA are between 1,400 and 1,500 bases long. In the present study we have used the same probe to identify transformants containing recombinant plasmids with inserts of human proenkephalin cDNA. In this manner we have been able to derive the nucleotide sequence of DNA complementary to the entire human proenkephalin mRNA sequence.

Isolation of cDNA clones

Poly(A) RNA was isolated from human pheochromocytoma tissue as previously described¹². Northern blot analysis of this

RNA, using an enkephalin pentadecamer pool (5'...CAT^AAAGCCGCC^ATA...3') containing two possible mismatches with the enkephalin-coding mRNA sequence, revealed that proenkephalin mRNA constitutes ~0.1% of the total poly(A) RNA from this tissue¹². A four- to fivefold enrichment of enkephalin mRNA was achieved by density gradient sedimentation (Fig. 2). A clone bank of 15,000 tetracycline-resistant transformants was made from cDNA copies of the enriched enkephalin mRNA.

To screen and maintain the pheochromocytoma clone bank, 7,000 transformants were picked and grown in 96-well microtitre plates. All cultures were transferred in ordered arrays to nitrocellulose for *in situ* hybridization with ³²P-labelled Met-enkephalin pentadecamer probe pool. Seventeen colonies hybridized strongly with the pool (Fig. 3). Plasmid DNA was isolated from the corresponding cultures and the size of cDNA inserts examined after endo *Pst*I cleavage by polyacrylamide gel electrophoresis. Most of the plasmids were found to contain both a 410 and a 600 base pair (bp) fragment. Several plasmids contained a third fragment of varying size. Two of the latter plasmids with the biggest inserts were selected for nucleotide sequencing together with the 600 and 410 bp fragments.

Sequencing strategy

To determine the nucleotide sequence of the cDNA inserts, recombinant plasmid DNA was cleaved with endo *Pst*I and the DNA fragments were cloned into the *Pst*I site of the bacteriophage vector M13mp7 RF-DNA. The single-stranded DNA prepared from recombinant phage served as a template for enzymatic sequencing reactions in the presence of chain terminators¹³ using a synthetic oligomer as universal primer¹⁴. Figure 4 shows the entire nucleotide sequence of preproenkephalin mRNA as derived from sequencing the cDNA inserts of two separate clones. One insert contained the entire 5' portion of the mRNA and ended ~20 nucleotides after the termination codon. The other had the entire 3' end but did not

	1	2	3	4	5	6	7	8	9	10	11	12	13
β -endorphin β (61-91)LPH	Tyr	Gly	Gly	Phe	Met		Thr	Ser	Glu	Lys	Ser	Gln	Thr
Met-enkephalin	Tyr	Gly	Gly	Phe	Met								
Leu-enkephalin	Tyr	Gly	Gly	Phe	Leu								
Dynorphin	Tyr	Gly	Gly	Phe	Leu	Arg	Arg	Ile	Arg	Pro	Lys	Leu	Lys
α -neo-endorphin	Tyr	Gly	Gly	Phe	Leu	Arg	Lys	Pro	(Gly, Tyr ² , Lys ² , Arg)				

Fig. 1 Structure of the opioid peptides. Only a partial structure of β -endorphin is shown. LPH, lipotropin.

contain the region corresponding to the 5'-untranslated region of the mRNA and the portion coding for the signal peptide (leader or presequence).

Primary structure of enkephalin precursor mRNA

The mRNA coding for Met- and Leu-enkephalin is 1,350 nucleotides long (Fig. 4), close to the size of 1,400 nucleotides observed in Fig. 2b. The 5'-untranslated region is remarkable for being >200 nucleotides long. We do not know how many nucleotides are missing from the capped 5' end of the mRNA. As artefacts can occur in the cDNA corresponding to the 5' end of the mRNA, we cannot rule out that the actual mRNA sequence is different in this region¹⁵. The 5' end features several stretches of homo-oligomeric pyrimidines, especially runs of UMP. The coding region starts with the first AUG and consists of 267 triplets with seven interspersed small repeat units corresponding to the enkephalin sequences. The 3'-untranslated region is 320 nucleotides long and is rich in UMP with several homo-oligomeric pyrimidine stretches towards the 5' end. An AAUAAA sequence common to most mRNA species¹⁶ is found 19 nucleotides before the poly(A) tail. It is curious that another such sequence occurs 20 nucleotides upstream. With both sequences represented in the chromosomal gene, it is possible that two populations of mRNA exist which contain either one or both sequences.

Primary structure of enkephalin precursor protein

The first 24 amino acids of the precursor protein have the characteristics of a leader sequence containing many hydrophobic amino acids, particularly Leu (Fig. 4). As Ala is the most

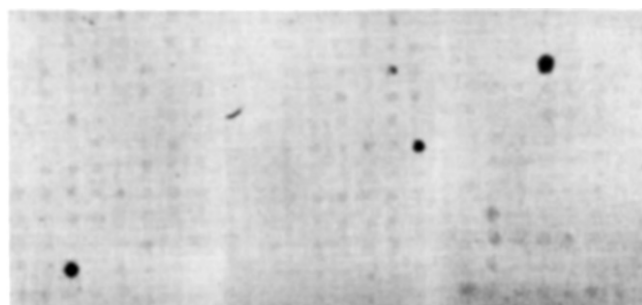


Fig. 3 *In situ* colony screening of a human pheochromocytoma cDNA library using ³²P-labelled enkephalin pentadecamer pool. Sucrose gradient fractions containing enkephalin mRNA (Fig. 2, lane 2) were pooled and used to direct the synthesis of single-stranded cDNA using reverse transcriptase and oligo(dT) as primer. Double-stranded cDNA was prepared using the large fragment of DNA polymerase I (Boehringer Mannheim), and the hairpin loops were cleaved with S₁ nuclease (Miles). Double-stranded S₁-treated cDNA was fractionated on a 10% polyacrylamide gel and cDNA 600–2,000 bases long was eluted. Forty ng of size-fractionated cDNA were obtained from 2 µg of enriched poly(A) RNA. Poly(dC) homopolymeric extensions were added to the 3' end using terminal transferase (PL Biochemicals). A clone bank of human pheochromocytoma cDNA was prepared by annealing the poly(dC)-tailed cDNA to endo PstI-cleaved plasmid pBR322 DNA tailed with dGMP²⁶, and the mixture used to transform competent *Escherichia coli* 294 cells. Approximately 15,000 tetracycline-resistant transformants were obtained from which <3% were ampicillin resistant. Seven thousand tetracycline-resistant colonies were picked and grown in 96-well microtitre plates. Colonies were transferred from the microtitre plates to nitrocellulose filters placed on 15 cm LB Petri plates containing 5 µg ml⁻¹ tetracycline using a 96-prong stainless-steel stamp. Colonies were grown on the filters for 6 h at 37 °C after which the filters were transferred to Petri plates containing LB with 12.5 µg ml⁻¹ chloramphenicol, and amplified overnight. Filters were lysed, neutralized and prepared for hybridization as described in ref. 27. Prehybridization of filters was for 6 h at 30 °C in 4× SSC, 2× Denhardt's, 40 mM NaPO₄, pH 7.8, and 300 µg ml⁻¹ heat-denatured herring sperm DNA. The pentadecamers were phosphorylated using T₄ polynucleotide kinase (New England Biolabs) and (γ-³²P)ATP (NEN) to a specific activity of 2 × 10⁶ c.p.m. pmol⁻¹. ³²P-labelled probe pool was purified by chromatography on 10 ml Sephadex G-50 columns. Each 15 × 7 cm filter was hybridized in 2 ml of the above hybridization mixture at 30 °C for 20 h containing 4 × 10⁶ c.p.m. pentadecamers. Filters were washed for 15 min, once at 25 °C and twice at 30 °C, in 4× SSC, 0.1% SDS, dried and exposed to X-ray film. Autoradiography was at -70 °C for 2 h using intensifier screens. Results are shown above for several filters containing positive signals.

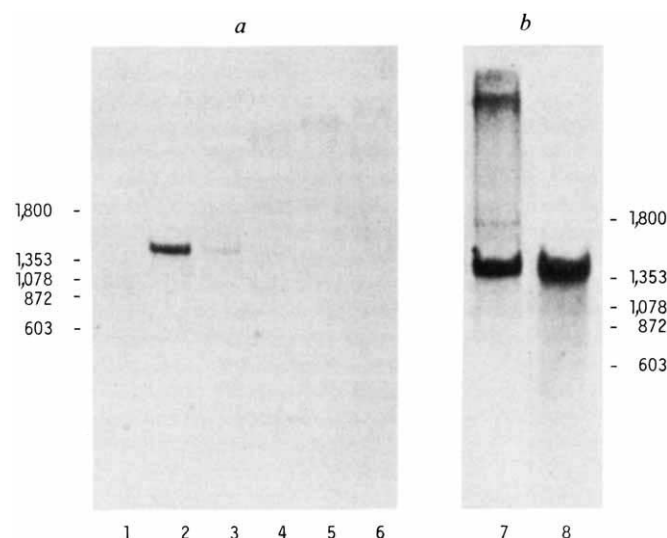


Fig. 2 Northern hybridization analysis of human pheochromocytoma poly(A) RNA using enkephalin probes. *a*, Fifty µg of human pheochromocytoma poly(A) RNA was fractionated on a 5–35% linear sucrose gradient at 4 °C for 16 h in a Beckman SW41 rotor at 35,000 r.p.m. Fractions from the gradient were precipitated and washed with EtOH and the pellets were resuspended in 10 µl of H₂O. Two µl aliquots from alternate fractions (0.8 µg) were denatured with glyoxal and electrophoresed in a 1.75% agarose gel for 3 h at 100 V. RNA was transferred from the gel to nitrocellulose²⁵. The RNA blot was hybridized with 6 × 10⁷ c.p.m. of ³²P-labelled pentadecamer probe pool (5 × 10⁶ c.p.m. pmol⁻¹) in 5 ml of 4× SSC, 1× Denhardt's, 0.1% SDS and 100 µg ml⁻¹ yeast tRNA at 30 °C for 2 days. Hybridization and washing of blots was done as described previously¹². Autoradiography was for 16 h at -70 °C using Dupont lightning plus intensifying screens. The top of the gradient is at the left hand side of the chart. *b*, Five µg of human pheochromocytoma poly(A) RNA was hybridized with 1 × 10⁷ c.p.m. pentadecamer probe pool (5 × 10⁶ c.p.m. pmol⁻¹; lane 7) and 3 × 10⁶ c.p.m. of nick translated pHPE 41 containing a nearly full-length proenkephalin cDNA insert (4 × 10⁸ c.p.m. µg⁻¹; lane 8). The length scale (number of nucleotides) was constructed by running glyoxal-denatured DNA fragments of *Hae*III digests of phase ΦX174 RF-DNA and *Hind*III digests of phage λ DNA on adjacent lanes of the gel.

common amino acid preceding the cleavage site of leader sequences¹⁷, we tentatively conclude that the leader sequence ends after Ala 24, as indicated in Fig. 5, although Ala 20 is another possible termination site. The N-terminal one-third of the mature proenkephalin molecule does not contain any enkephalin sequences but consists of a 73 amino acid peptide (residues 25–97 in Fig. 4) containing all six Cys residues in the protein. A proper folding for efficient processing is probably ensured by the formation of disulphide bridges in this part of the precursor.

It is tempting to speculate that the occurrence of multiple enkephalin sequences in the same molecule is due to duplication events of one small genetic unit in the distant past. If so, sequence changes dictated by diversified functional and processing requirements may have erased most of the supporting evidence, as can be seen by analysing the structure of the regions that separate the enkephalin units. The length of these 'spacer' regions ranges from 2 amino acids (between units 1 and 2) to 41 (between units 3 and 4). Four of these spacer regions contain highly hydrophilic sequences with several Glu and Asp residues in a row. The spacer peptide that separates Met-enkephalin unit 5 from Leu-enkephalin is missing this hydrophilic feature. It is interesting to note that this peptide (peptide E) is totally conserved in the bovine and human enkephalin precursors and is a more potent opioid agent than the enkephalins¹⁰. Perhaps processing of the hydrophobic spacer is less efficient than processing of the hydrophilic spacers. The hydrophobic spacer might ensure that peptide E is released intact from the precursor to serve as an opioid agent or another kind of regulator.

Fig. 4 Primary structure of human Met- and Leu-enkephalin in precursor mRNA and protein. Plasmid DNA was isolated by mini-screen procedure from 5 ml of overnight cultures and digested to completion with endo *Pst*I (New England Biolabs). Digested DNA was ligated to 10 ng of *Pst*I-cleaved M13mp7 RF-DNA in 20 μ l containing 50 mM Tris-HCl, pH 8, 0.1 mM EDTA, 7 mM MgCl₂, 10 mM dithiothreitol, 500 μ M rATP and 0.2 U of T₄ DNA ligase (New England Biolabs) for 1 h at room temperature. The mixture was used to transform competent *E. coli* JM103 cells and transformed cells were plated with top agar in the presence of isopropyl-tgio- β -D-galactoside and 5-bromo-4-chloro-3-indolyl- β -D-galactoside. After a 10 h incubation at 37°C, colourless plaques were transferred to 1.5 ml 2 $\times$ yeast-tryptone broth. Bacteria were sedimented from grown cultures by centrifugation and phage were precipitated from the supernatant with polyethylene glycol-6000 (3%) and NaCl (0.5 M). DNA was isolated by phenol-CHCl₃ extraction, precipitated with ethanol and redissolved in 10 μ l of H₂O. DNA templates were sorted by single-lane tracking to avoid redundancy and non-productive sequencing generated by the empty phage DNA or poly(dG)-tailed sequences at the insertion site. For the final sequence determination, six templates were chosen which originated from two different cDNA clones. Sequencing reactions were done for 15 min at 25°C in 4 μ l of 10 mM Tris, pH 8, 0.1 mM EDTA, 7 mM MgCl₂, and 1 mM dithiothreitol containing about 100 ng of template DNA, 0.2 pmol of primer, 0.2 U *E. coli* DNA Polymerase I (Klenow fragment; Boehringer Mannheim), 5 μ Ci of [α -³²P]dCTP (Amersham, 400 Ci mmol⁻¹) and deoxy- and dideoxynucleoside triphosphates in the proper concentrations. A chase was performed for 15 min by adding 1 μ l of a 2 mM solution of each of the deoxyribonucleoside triphosphates and reactions were stopped by adding 10 μ l of a formamide-dye mixture. After heating the reaction mixtures to 100°C for 3 min, 1 μ l aliquots were electrophoresed on thin 40 cm long 5% polyacrylamide-8 M urea gels for 2-6 h at 25 mA. Gels were dried onto Whatman 3 mm paper and exposed to X-ray film overnight.

Special features of the human enkephalin precursor and possible cleavage products

Some of the special features of the enkephalin precursor are presented in Fig. 5. Five of the seven enkephalin moieties are flanked on both sides by pairs of basic amino acid residues (enkephalin moieties 1-3, 5 and 6, Figs 4 and 5), suggesting that they are cleaved out of the precursor by the combined action of a trypsin-like enzyme and a carboxypeptidase. Action of these

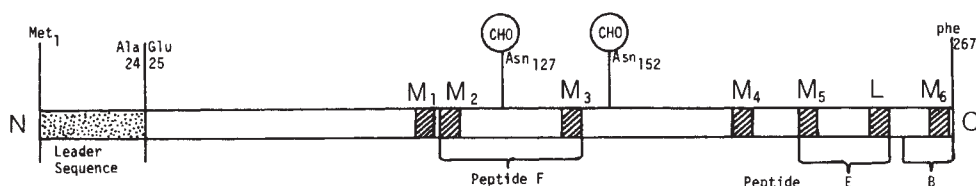


Fig. 5 Low-resolution model of enkephalin precursor showing the distribution of Met-enkephalin sequences (M_1 , M_2 , M_3 and so on) and Leu-enkephalin (L). The position of potential carbohydrate (CHO) attachment sites and the sequences corresponding to peptides F, E and B of bovine adrenal medulla¹⁰ are depicted.

enzymes would also be expected to generate Met-enkephalin-Arg-Gly-Leu from enkephalin sequence 4 and Met-enkephalin-Arg-Phe from the carboxy-terminal region of the protein. Met-enkephalin-Arg-Phe has actually been observed in bovine adrenal medulla⁹.

The biologically active domains of other prohormones, such as proinsulin²¹ and POMC²², are also flanked on both sides by pairs of basic amino acid residues, suggesting that similar trypsin-like enzymes are involved in processing of these prohormones. The amino acid sequence in Fig. 4 also reveals the existence of two potential glycosylation sites in the precursor at Asn 127 and Asn 152. The sequence Asn-X-Ser or Asn-X-Thr is required to attach carbohydrate to an Asn in a protein²³. However, as not all Asn residues in such a sequence are glycosylated, it is not possible to predict whether the enkephalin precursor contains Asn-linked oligosaccharides²³.

Comparison of human pro-enkephalin- and bovine enkephalin-containing peptides

Kilpatrick *et al.*¹⁰ have sequenced several large enkephalin-containing peptides isolated from chromaffin granules of bovine adrenal medulla. Comparison of the sequence of three of these peptides with the sequence of human proenkephalin shows that all three are contained within the human precursor. The sequence of bovine peptide B corresponds to amino acid residues 238–267 (the C terminus of the protein) of the human precursor with at least one amino acid substitution at residue 240 (a Pro in the bovine peptide for Ala in the human)¹⁰. The sequence of peptide E (a more potent opioid than the enkephalins)¹⁰, which contains both a Met- and a Leu-enkephalin moiety, corresponds exactly to the sequence of residues 210–235 in the human enkephalin precursor. The sequence of peptide F (34 amino acids long) corresponds to the sequence of residues 107–140 in the human precursor with five amino acid substitutions. The amino acid substitutions in human peptide F are (human for bovine) in positions 120 (Met for Leu), 122 (Pro for Val), 129 (Ser for Gly), 131 (Leu for Val) and 133 (Ala for Gly). It is interesting that peptides E and F have pairs of basic amino acid residues flanking both sides of the enkephalin sequences they contain. As pairs of basic amino acid residues are potential tryptic cleavage sites, this finding suggests that peptides E and F are processing intermediates in the formation of enkephalins. Yet the high potency of peptide E as an opiate and the complete conservation of sequence between human and

bovine peptides indicate that this peptide is an important opiate compound in itself.

Enkephalin-like immunoreactivity has been observed not only in several areas of the brain but also in sympathetic ganglia, nerve plexuses in the intestine and nerve fibres in adrenal medulla and pheochromocytoma. It would be very interesting to know whether the same array of opioid peptides is produced in these tissues as in adrenal medulla. Tissue-specific use of enkephalin peptides could occur by differential processing of the precursor molecule or by selective expression of one or more of a family of enkephalin genes in different tissues. Isolation and characterization of enkephalin genes will be an important step in attempting to answer this question.

Close examination of the structure of the human enkephalin precursor reveals that it does not contain the sequence of any of the other opioid peptides, β -endorphin, dynorphin or α -neo-endorphin. Hence, at least three of the four types of opioid peptides are derived from different primary gene products.

The complete sequences of the precursors to β -endorphin (POMC) and enkephalins are now known. In both precursors there is repetition of a sequence that is flanked on both sides by pairs of basic amino acid residues. POMC has three copies of the MSH sequence²² and the enkephalin precursor has seven copies of the enkephalin sequence. J. Kurjan and I. Herskowitz have recently discovered that α -factor, a peptide pheromone 13 amino acids long that promotes sexual union of yeast gametes, is synthesized from a precursor containing four copies of the peptide in an evenly spaced arrangement (personal communication). The α -factor sequences are flanked on only one side by a pair of basic amino acid residues indicating that the proteolytic processing of the precursor is different from that of proenkephalin and POMC.

Thus widely diverse organisms make use of the multivalent precursor for the synthesis of important regulatory peptides. In the case of POMC, peptides from different regions of the precursor (MSH, ACTH and endorphin) may act as agonists or antagonists in coordinating a behavioural response such as the response to stress²⁴. The function of the multi-enkephalin precursor is less clear.

This work was supported by National Institutes of Health Grants AM-16879 and GM07759. We also thank R. Swanson for his support.

Note added in proof: While this paper was in press two papers appeared in the 21 January issue of *Nature* on the structure of the enkephalin precursor in bovine adrenal gland^{28,29}. The distribution of enkephalin sequences in the human precursor is very similar to that in the bovine precursor.

Received 28 December 1981; accepted 15 January 1982.

- Hughes, J. *et al.* *Nature* **258**, 577–579 (1975).
- Li, C. H. & Chung, D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1145–1148 (1976).
- Bradbury, A. F., Smyth, P. G., Snell, C. R., Birdsall, N. J. M. & Hulme, E. C. *Nature* **260**, 793–795 (1976).
- Goldstein, A., Tachibay, A. S., Lowney, L. I., Hunkapiller, R. M. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6666–6670 (1979).
- Kangawa, K., Matsuo, H. & Igarashi, M. *Biochem. biophys. Res. Commun.* **86**, 153–160 (1979).
- Mains, R. E., Eipper, B. A. & Ling, N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3014–3018 (1977).
- Roberts, J. L. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4826–4830 (1977).
- Roberts, J. L. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5300–5304 (1977).
- Stern, A. S. *et al.* *Biosynthesis, Modification and Processing of Cellular and Viral Proteins* (eds Koch, G. & Richter, D. (Academic, New York, 1980).
- Kilpatrick, D. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 3265–3268 (1981).
- Giraud, P. & Eiden, L. E. *Biochem. biophys. Res. Commun.* **99**, 969–975 (1981).
- Comb, M. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Sanger, F., Nickle, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).

- Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309–321 (1981).
- Richard, R. I., Shine, J., Ullrich, A., Wells, J. R. E. & Goodman, H. M. *Nucleic Acids Res.* **7**, 1137–1146 (1979).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).
- Steiner, D. F., Quinn, P. S., Shu, J. C., March, J. & Tager, H. S. *Ann. N.Y. Acad. Sci.* **343**, 1–16 (1980).
- Dandekar, S. & Sabol, S. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Lewis, R. V., Stern, A. S., Rossier, J., Stern, S. & Udenfriend, S. *Science* **208**, 1459–1561 (1980).
- Beaumont, A., Feuntes, J. A., Hughes, J. & Metters, K. M. *FEBS Lett.* **122**, 135–137 (1980).
- Steiner, D. F., Cunningham, D., Spiegelman, L. & Aton, D. *Science* **157**, 697–701 (1967).
- Nakanishi, S. *et al.* *Nature* **278**, 423–427 (1979).
- Marshall, R. D. A. *Rev. Biochem.* **41**, 673–702 (1972).
- Pedersen, R. C. & Brownie, A. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2239–2243 (1980).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
- Deng, G. & Wu, R. *Nucleic Acids Res.* **9**, 4173–4188 (1981).
- Grinstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
- Noda, M. *et al.* *Nature* **295**, 202–206 (1982).
- Gubler, U., Seeburg, P., Hoffman, B. J., Gage, L. P. & Udenfriend, S. *Nature* **295**, 206–209 (1982).

Studies of X chromosome DNA methylation in normal human cells

Stanley F. Wolf & Barbara R. Migeon

Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Studies of methylation along 28 kilobases of X-chromosome DNA, assayed by Southern blot analysis using cloned X-chromosome-specific probes, indicates that X DNA methylation in normal human cells changes with replication, is not correlated with number of X chromosomes or transcriptional activity, and is less stable and more prevalent than when the human X is in the foreign environment of mouse-human hybrid cells. In contrast with observations of others in heteroploid cells, we observed no derepression of the inactive X in clonal populations of normal human fibroblasts treated with 5-azacytidine. This may reflect the differences in stability of the methylation of the human X in a foreign environment. Our observations preclude ubiquitous methylation differences as the molecular basis for X-chromosome inactivation.

DNA METHYLATION has frequently been proposed to be responsible for the regulation of developmental processes¹⁻³. These differentiation events include that resulting in the inactivation of all but one X chromosome in somatic cells of mammals⁴. Riggs³ proposed that DNA methylation was responsible for initiation and maintenance of mammalian X inactivation, invoking the presence of sequence methylases with preference for half-methylated sites. Methylation has been implicated in X-chromosome inactivation on the basis of indirect evidence: the ubiquity of 5-methylcytosine, the existence of sequence-specific methylases and tissue-specific differences in methylation of differentiated genes^{5,6}. There is some evidence that a gene on the inactive X remains incapable of expression even when protein superstructures are removed⁷, implying that the DNA of the inactive X has undergone a permanent modification. The tissue-specific association of undermethylation with gene activity⁸⁻¹⁷ and abundant methylation in the early embryo¹⁵ has led to speculation that the inactive X chromosome might be highly methylated^{14,15}. Recent studies of interspecies cell hybrids have shown that 5-azacytidine (5-azaC) can induce local derepression of the inactive X chromosome, presumably because the cytidine analogue has blocked methylation at relevant sites on the inactive X¹⁸.

We have identified two cloned DNA fragments unique to the X chromosome¹⁹ that provide a more direct means of looking for methylation differences between inactive and active X chromosomes. Using these probes, we have compared DNA methylation in placentas and skin fibroblasts from males and females. Furthermore, we have carried out direct experiments testing the ability of 5-azaC to induce reactivation of the silent X in normal human cells having enzyme markers that distinguish active from inactive X. Our results show no significant sex differences in methylation of X sequences. Moreover, studies of clonal populations of normal human cells demonstrate that the pattern of DNA methylation is not faithfully transmitted.

X-chromosome DNA methylation studies

X-chromosome-specific DNA probes provide a direct means of testing the hypothesis that the inactive X chromosome is highly methylated. The restriction enzymes *HpaII* and *MspI* recognize the sequence CCGG but *HpaII* cannot cut if the internal C is methylated²⁰. If DNA methylation of active and inactive X chromosomes differs significantly, the restriction fragments generated by digestion with *HpaII* and hybridization with a labelled X-specific probe should differ between DNA of male cells with a single active X and that of female cells that have an inactive X as well.

The plasmids 0307 and 0604 contain sequences specific to the human X chromosome¹⁹. By means of mouse-human hybrids, the insert segments in the 0307 and 0604 plasmids have been mapped to the region on the long arm between Xq13 and Xq25. Mapping was carried out using a series of hybrids derived from mouse cells (Sv-tfm or A9) and human cells with balanced X/autosome translocation that have been used to map other X-linked genes^{21,22}. The 0307 plasmid contains DNA that is transcribed as it hybridizes to total RNA extracted from normal human fibroblasts and this hybridization is RNase sensitive (Fig. 1). The 3.4-kilobase (kb) insert is a single copy sequence based on (1) analysis of reassociating kinetics using human genomic DNA¹⁹, (2) intensity of hybridization to genomic DNA in Southern blots, relative to that obtained with a γ -globin probe¹⁹ and (3) identification of a *PstI* polymorphism within the sequence that behaves as expected for a single copy sequence. (Southern blot analysis of human lymphocyte DNA digested with *PstI* and probed with 0307 usually results in bands at 7.4 and 1.0 kb. In one male with a *PstI* restriction site polymorphism, the 7.4-kb band has been replaced by a 4.2-kb band. If the 0307 sequence were present more than once on the X, one would expect a new band in addition to the original ones.)

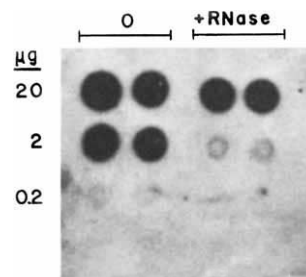


Fig. 1 Hybridization of 0307 DNA to cultured human fibroblast RNA. Total RNA, purified using hot phenol-SDS³⁰, was aliquoted precisely into two tubes. Pancreatic RNase A (Sigma: 5 × crystallized, boiled 10 min) was added to one tube (100 µg ml⁻¹) and incubated (37 °C, 2 h); the second tube was untreated. The RNA in serial dilutions was dotted onto nitrocellulose filters¹⁶ and the filter was hybridized to 0307 labelled with ³²P by nick translation¹⁹. The final wash was at 60 °C, 0.1 SET and exposed to Kodak XAR5 X-ray film (-70 °C, 6 days). Hybridization did not occur in the absence of RNA. Parallel filters hybridized with γ -globin cDNA probe showed hybridization only to the 20-µg dots and no RNase sensitivity.

Because the X-chromosome DNA sequence in the plasmid 0307 contains no *HpaII* site, 0307 is a probe for methylation of *HpaII* sites in neighbouring DNA. The 7.5-kb insert in plasmid 0604 includes a single *HpaII* site.

Methylation patterns in male and female placentas and uncloned fibroblast cultures

Figure 2 shows the results of *HpaII* and *MspI* analysis of placentas (chorionic villi) and uncloned fibroblasts using probe 0307. The probe hybridizes to a single band (4.6 kb) in each of the DNA samples digested with *MspI* (Fig. 2, lanes 1–6) so that the distance between *HpaII* sites which span the 0307 sequence is the same in all cells. In *HpaII* digests of male and female specimens (Fig. 2, lanes 7–12) the 0307 probe hybridizes to multiple fragments, ≥ 4.6 kb. Hybridization to fragments larger than 4.6 kb indicates that the internal C in the *HpaII* recognition sequence is methylated in site(s) flanking the 0307 sequence.

The multiple bands observed in *HpaII* digests probed with 0307 are unlikely to be the result of partial digestion. Samples were digested with excess enzyme for up to 20 h (see figure legends for details). Moreover, the number and relative intensity of bands, although variable in DNA from different individuals, are highly reproducible in repeat assays of DNA from the same individual. Furthermore, a recombinant λ phage with a 15-kb human DNA insert that includes the 0307 sequence has the *HpaII* sites necessary to produce the multiple bands we observe. (We have identified and purified a recombinant phage contain-

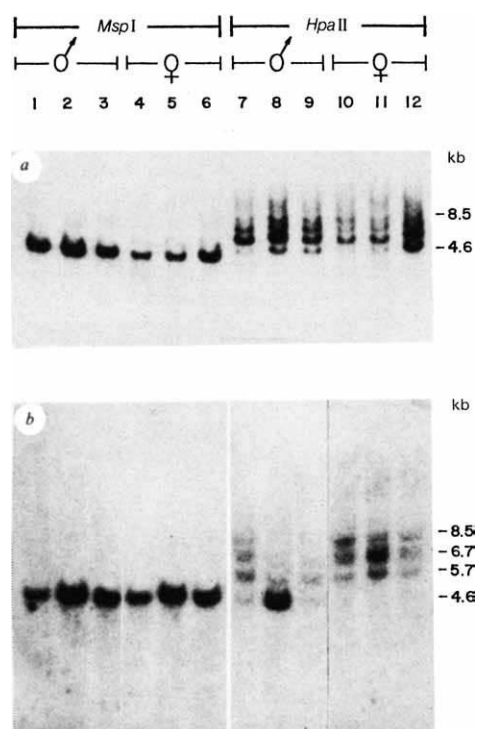


Fig. 2 Comparison of X DNA methylation in male and female cells. High molecular weight DNA was purified from chorionic villi of human placentas (a) and cultured human diploid fibroblasts (b). Southern blots³¹ of enzyme-digested DNA, fractionated on 0.8% agarose gels, were hybridized with the X-specific DNA probe 0307. Digestions were carried out for 16 h with 24 U *MspI* or *HpaII* per μ g of DNA in conditions recommended by the manufacturer. Filters washed at 65 °C in 0.1 SET were exposed to X-ray film (–70 °C, 5 days).

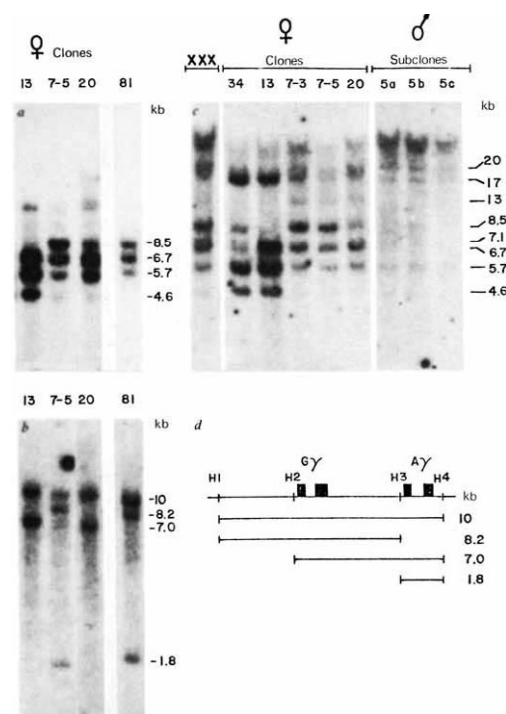


Fig. 3 DNA methylation in cloned human fibroblasts. a, DNA of fibroblast clones independently isolated from a culture of 46XX female cells (see text) was digested with *HpaII* (20 U per μ g DNA, 16 h). The filter was hybridized with ³²P-labelled 0307, washed at 55 °C in 0.1 SET and exposed to X-ray film (–70 °C, 5 days). b, Hybridized probe was removed from the filter used in a by washing (70 °C, 70% formamide, 0.1 SET). The filter was hybridized with ³²P-labelled γ -globin cDNA insert of plasmid JW151³³. After a final wash (65 °C, 0.1 SET), the filter was exposed to X-ray film (–70 °C, 12 days). c, DNA from clones of female fibroblasts and subclones of male origin (son of the source of female clones) was digested with *HpaII* (10 U per μ g DNA for 16 h when a further 3 U enzyme per μ g DNA was added and the incubation continued for 4 h). Electrophoresis was longer than in Figs 1 or 2a and DNA was transferred for 48 h to enhance the resolution and transfer of high molecular weight fragments. The filter was hybridized with the 0307 X DNA probe, washed (55 °C, 0.1 SET) and exposed to X-ray film (–70 °C, 5 days). Lane 1, DNA from 47XXX fibroblasts; lanes 2–4, 6, DNA from clones of the same female as in a; lane 5, DNA from the same clone as in lane 4, purified after five further population doublings; lanes 7–9, independent subclones derived from a single clone of male fibroblasts. d, *HpaII* restriction site map of human γ -globin loci¹¹.

ing the 0307 sequence from a *HaeIII*-*AluI* human genomic DNA library²⁴. Southern blots of the recombinant phage DNA, partially digested with *MspI* and probed with 0307, show the same 4.6–13-kb bands seen in blots of genomic DNA digested with *HpaII*.)

Because the 0307 sequence is single copy, yet hybridizes to multiple restriction fragments in *HpaII* digests, the region around the 0307 sequence must be methylated differently from cell to cell. Similar studies with the second X-specific cloned DNA (0604) as probe indicate that the methylation of the single internal *HpaII* site is also variable in both male and female fibroblasts (data not shown). The complexity of the banding pattern is not obviously related to the number of X chromosomes in the cell, as some cultures from females have as many bands as those found in trisomic cells with three X chromosomes (including two inactive ones) (Fig. 3a, lane 1).

Therefore, DNA methylation in the regions probed by both X-specific sequences is characterized by variability and we observe no consistent sex differences in placentas or fibroblasts.

DNA methylation in clonal populations of fibroblasts

The multiple bands we observed in chorionic villi and uncloned fibroblast cultures might be attributed to a mixed population of cells, each with a fixed but different methylation pattern. To test this hypothesis, we examined methylation in clonal populations derived from dermal fibroblasts of a female heterozygous for the X-linked marker glucose 6-phosphate dehydrogenase (G6PD) A. Clones were obtained by plating 10 cells per 60-mm dish and clonal cultures established by picking well isolated colonies, using cloning cylinders²⁵ to ensure that all cells in each clonal culture were progeny of a single cell. Furthermore, purity could be monitored on the basis that each clone, although of heterozygous genotype, expressed only a single G6PD isozyme, the one coded by the active X chromosome⁴. DNA was also purified from clonal populations of fibroblasts from a male, as well as derivative subclones to ensure that each colony was of homogeneous composition. The Southern blots of DNA from these clonal cultures generated by *Hpa*II digestion and hybridization with the 0307 probe show multiple bands (Fig. 3a, c) similar to those observed in uncloned cultures.

Hybridization to more than one restriction fragment in DNA from clonal populations of male cells or more than two in DNA from clones of female cells indicates that during the 20 generations required for these clonal populations to proliferate, the methylation of *Hpa*II sites in the DNA adjacent to the 0307 sequence was not fixed. Methylation in the vicinity of the γ -globin loci in the same clones is also heterogeneous, as at least two patterns were observed when the same filters were hybridized with a γ -globin cDNA probe (Fig. 3b).

Although the methylation of *Hpa*II sites in X DNA differs among progeny of single cells, the relative intensities of the various *Hpa*II bands generated indicate that the changes are not random. No clone produced all possible bands (compare uncloned fibroblasts in Fig. 3c, lane 1, with others). Moreover, the banding pattern (relative intensity and size of restriction fragments) for each clone is characteristic of that clone and is relatively stable. For example, the pattern of multiple bands produced by clone 7 remained the same after six further cell doublings (Fig. 3c, lanes 4, 5).

Methylation of human X chromosomes in male subclones

The difference in the banding patterns between various clones of female fibroblasts more probably reflects the unique origin of each clone. We know that clones 13 and 20 originated from different skin fibroblasts of this female because each expresses a different G6PD allele, and the other clones, as they were derived from a heterogeneous uncloned culture, may also have a unique origin.

On the other hand, subclones obtained from the homogeneous cloned population in the male are all progeny of the same cell. Therefore, our observation that three of these subclones have restriction fragments of identical size and relative intensity (Fig. 3c, lanes 7–9) indicates that the pattern as a whole is conserved over at least 20 population doublings. It seems that methylation of individual *Hpa*II sites is not stably inherited, but the probability that a site will be methylated is heritable.

Methylation of human X chromosomes in mouse-human hybrid cells

Although the methylation at *Hpa*II sites around the 0307 sequence is not fixed in cultured human diploid fibroblasts,

methylation at these sites is more stable when the human X chromosome is present in a mouse-human hybrid cell. Figure 4 shows the Southern blot analysis of DNA from four independently derived hybrid clones (each with a different active human X) digested with *Hpa*II and probed with 0307.

The hybrid clones produced bands of the same size as those in DNA from normal human male cells, so that the sites of methylation are the same for active X DNA in either cell. However, the striking predominance of one band in each hybrid clone is evidence for greater stability of methylation of the human X in the foreign environment of the mouse-human hybrid, than in its normal habitat.

It is notable that none of the four mouse-human hybrid cells generated bands >6.7 kb. While the absence of longer bands could be merely fortuitous, it may be that the human X chromosome in hybrid cells is in fact less methylated than in human cells.

Relationship of methylation to X-chromosome activity

One might expect that the banding pattern would be identical for clones of cells derived from the same culture, each with the same active X chromosome. However, clones 7, 13 and 34 from cultures of a heterozygous female have the same active X (maternal X), as they express the G6PD A allele, but have different banding patterns; patterns also differ for clones 81 and 20, which express the G6PD B allele (paternal X is active; Fig. 3a, c).

The presence of both active and inactive X chromosomes in these female cells may obscure specific banding patterns associated with the functional state of the chromosome. It is clear, however, that female cells do not have large *Hpa*II X DNA fragments that are not present in males. In fact, the single active X chromosome in male fibroblasts can be highly methylated in the 0307 region, as indicated by the presence of very large fragments in *Hpa*II-digested DNA from some male cells (Fig. 3c, lanes 7–9). However, the same sites in X chromosomes of other male cells are not methylated (Figs 2, 3c, lanes 7–9). The predominant band from the single active human X in various mouse-human hybrid cells can also differ in size, but tends to be of lower molecular weight, compatible with less methylation of these active X chromosomes (Fig. 4). Thus, no consistent methylation differences can be seen with respect to X-chromosome activity.

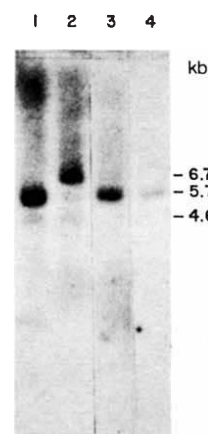


Fig. 4 Methylation in the single active human X chromosome of mouse-human hybrid cells. Purified DNA was digested with *Hpa*II (20 U per μ g, 16 h). Hybridization was with labelled 0307 X DNA probe. Final wash: 55 °C, 0.1 SET; exposure: –70 °C, 5 days. Lanes 1, Sv-tfm $\times$ Kop (clone 4a)²²; 2, Sv-tfm $\times$ 297 (clone 1)²²; 3, A9 $\times$ Gm 1322 (clone 9-1b)¹⁹; 4, A9 $\times$ clone 81 (clone 6a).

Table 1 Effect of 5-azacytidine on X chromosomal phenotype

Expt	Cells per dish ($\times 10^4$)	5-azaC (μ M)	Exposure (h)	Recovery (days)	Cells analysed ($\times 10^5$)	HPRT phenotype HAT ^r colonies	Labelled cells	G6PD phenotype
1	20	3	24	5-6	20.0		0	
	20-40	3-10	24-48	1-14	120.0			A
2	0.4-2.0	2-5	68	0	1.1	0		
	14	5	68	6	20.0	0		
3	0.2-0.5	2-3	24	1-4	3.2	0		
4	4	2	48	0	20.0	1		A
	40	2	48	4	20.0	0		
	40	2	48	6	40.0		0	
	40	2	48	12-14	100.0			A

Effect of 5-azacytidine on X-chromosome activity and methylation

Studies were carried out to determine if 5-azaC could derepress the inactive X chromosome in normal human skin fibroblasts. The two X chromosomes in female cells used for this study differ with respect to alleles at two loci, being heterozygous for G6PD A as well as for hypoxanthine phosphoribosyl transferase (HPRT) deficiency (G6PD A/B; HPRT $-/+$). For these experiments we treated clonal populations of cells that expressed G6PD A and HPRT deficiency (the alleles coupled on the active X). The phenotypic changes expected if 5-azaC were effective in derepressing the X were the re-expression of G6PD B and the wild-type HPRT alleles on the inactive X chromosome.

Four experiments were carried out, essentially similar but differing in the conditions of exposure to 5-azaC (see Table 1). Cells were plated at densities of 2×10^3 – 2×10^5 cells per dish in normal growth medium (minimal essential medium + 15% fetal calf serum, Gibco), and treated 24 h later with a fresh solution of 1–10 μ M 5-azaC for 24–72 h. Cells were analysed for evidence of reactivation of the inactive X by: (1) hypoxanthine/aminopterin/thymidine (HAT) selection for cells that had acquired HPRT activity, (2) autoradiography of cells exposed to 3 H-hypoxanthine, or (3) G6PD electrophoresis to detect cells expressing the G6PD B allele on the inactive X.

More than 1.7×10^7 cells were treated with 5-azaC; of the 6.4×10^6 cells selected in HAT medium, we observed only one subclone that had undergone a phenotypic change. That colony (HAT^r-5-azaC-1) was able to proliferate in HAT medium, having acquired HPRT activity, but it expressed only G6PD A. Studies of mouse-human hybrids derived from this clone suggest that the change is at the HPRT locus on the active X chromosome (Fig. 5).

Although the lack of HAT^r subclones might be related to poor proliferation of cells in which derepression had occurred at the HPRT locus, the cloning efficiency for treated cells (Table 2) was sufficient to reveal reactivants if they occurred with 10^{-3} frequencies reported for mouse cells¹⁸.

To circumvent the problem of cell proliferation, a further 6×10^6 cells were examined by autoradiography. Cells in nonselective medium were exposed to 3 H-hypoxanthine (NEN, 3.6-10C/M, 10 μ Ci ml⁻¹ for 18 h); the dishes were then washed, fixed and treated with 5% trichloroacetic acid. In some cases, the cells were replated following recovery so that cell density in autoradiographs did not exceed 10^6 cells per dish. The dishes were emulsed in Kodak NTB-2, developed 1 week later, and screened microscopically for the presence of labelled cells. The assay is sensitive enough to detect individual cells that have acquired HPRT activity and only 10% normal activity is necessary for a cell to be labelled in the conditions used; however, no labelled cells were observed.

Finally, more than 20 cultures of cells totalling 3.6×10^7 cells were examined for G6PD phenotype on cellulose acetate²⁵ in conditions sensitive enough to detect an admixture of enzyme where the minor component is as little as 5%. However, the G6PD B allele on the inactive X was never observed.

The experimental conditions that were used induced a decrease in cloning efficiency (Table 2) similar to that reported for phenotypically altered mouse cells^{26,27}. Furthermore, we observed changes in methylation at CCGG sites with both X DNA and γ -globin probes (Fig. 6); namely, a relatively greater intensity of small *Hpa*II bands with 0307 probe and the appearance of a 5.2-kb band with γ -globin probe (see Fig. 3d). This hypomethylation, although more subtle than that reported from more indirect observations in mouse cells²⁷, is compatible with changes seen in retroviral sequences from transformed chicken cells¹⁶. We observed an effect of treatment only in DNA from cells treated with 2 μ M 5-azaC 7 days previously; none was apparent directly after treatment (Fig. 6) or with 5–10 μ M 5-azaC at any time (data not shown).

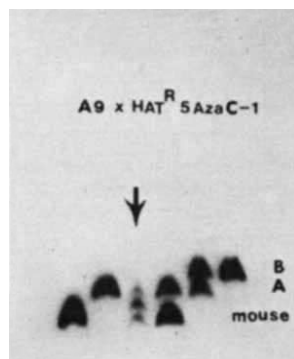
Therefore, despite the use of conditions able to induce methylation changes and toxicity, we have obtained no compelling evidence that 5-azaC can derepress HPRT or G6PD loci on the inactive X chromosome in normal human cells carrying suitable markers.

Implications

The rationale for implicating DNA methylation as a mechanism for regulating gene activity is based on the presumption that, as in bacterial systems, the methylation pattern of specific DNA sequences is stably transmitted from progenitor to progeny. Our studies of clonal populations of normal diploid fibroblasts demonstrate that, in fact, methylation of the *Hpa*II sites we examined is not stably transmitted in newly replicated X DNA. There is other evidence that methylation may not be stably inherited. Methylation of the chicken thymidine kinase gene changes during propagation of the recipient mouse cells¹⁴. In addition, 5-azaC-induced expression of the metallothionein I gene in mouse thymoma cells, attributed to undermethylation, is unstable within five cell generations¹⁷. In these cases, the instability might be attributed to the foreign environment, experimental manipulation or heteroploidy. However, our results indicate that methylation is unstable in entirely normal euploid cells.

Our observations are limited to an examination of *Hpa*II sites around the two X-chromosome DNA probes that have not been characterized with regard to translation. However, we have

Fig. 5 G6PD electrophoresis of mouse-human hybrid showing G6PD A and heterodimer (arrow). The 5-azaC-induced HAT^r clone was hybridized to A9 mouse cell and hybrids selected in HAT^r. Because the HPRT activity in clone HAT^r 5-azaC-1 was significantly less than that of a wild-type allele (10–20% of control), hybrids proliferated poorly and only one was able to produce sufficient cells for analysis. This hybrid clone expressed human G6PD A, mouse G6PD and the heterodimer, so that the chromosome contributing HPRT activity was that with the G6PD A allele and therefore the active X. Lanes: 1, A9; 2, HAT^r 5-azaC-1; 3, hybrid; 4, mixture of parental cells; 5, mixture of G6PD B and G6PD A fibroblasts; 6, G6PD B control.



explored a significant portion of the human X chromosome. Based on the total number of *HpaII* fragments and the size of the largest band (see Fig. 3c), we estimate that with probe 0307 we have examined the methylation of at least eight *HpaII* sites in a segment of ~20 kb in the middle of the long arm of the X. With probe 0604, we have examined another site within a 7.5-kb sequence at Xq13-25. Evidence from DNA transfer experiments and observations of local derepression at X loci suggest that modifications of DNA associated with inactivity are likely to be numerous and occur over relatively small regions of the chromosome^{7,28}.

Because there are no large 0307-specific *HpaII* fragments unique to female cells, we can exclude the possibility that ubiquitous methylation accounts for inactivity of the silent X. The failure of hypomethylation, induced by 5-azaC, to reactivate alleles on the inactive X chromosome provides further evidence that X-chromosome inactivity in normal human cells is not influenced by the state of methylation at all sites along the chromosome. It is conceivable that phenotypic changes observed in transformed mouse and chick cells treated with 5-azaC^{16-18,26,27,29} are not attributable to direct effects on DNA methylation. The HAT^r clone we obtained probably reflects reversion at the HPRT locus on the active X and other evidence that 5-azaC can induce mutation has been reported²⁹.

On the other hand, our observations do not preclude a relationship between X-chromosome methylation and transcriptional activity. The instability of methylation at individual *HpaII* sites and the different methylation patterns observed for

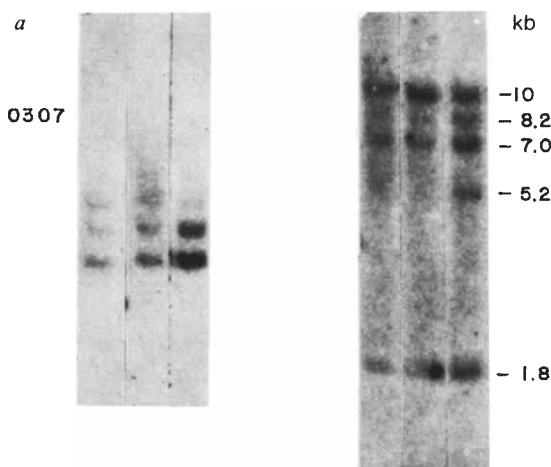


Fig. 6 Methylation of *HpaII* restriction sites in human diploid fibroblasts treated with 5-azaC. Lanes: 1, untreated; 2, 2 μ M, 48 h; 3, 2 μ M, 48 h, recovery 7 days. *a*, DNA was digested with *HpaII* (20 U per μ g DNA, 16 h) and Southern blots were hybridized to the labelled 0307 probe. Final wash: 55 °C, 0.1 SET; exposure: -70 °C, 4 days. *b*, after autoradiography, the X probe was removed and the filters rehybridized with a γ -globin cDNA probe as described in Fig. 3b legend. Exposure: -70 °C, 7 days.

Table 2 Effect of 5-azaC on proliferation of fibroblast clones (expt 4, Table 1)

5-azaC (μ M)	Length of exposure (h)	Clones per cells plated	Cloning efficiency	Relative cloning efficiency (%)
0		60/150	0.40	100
2	24	34/150	0.23	57.5
	48	41/150	0.27	67.6
5	24	14/150	0.09	22.5

the active X chromosome in male cells indicate that many arrays of methylation along the X chromosome are compatible with X-chromosome activity (some highly methylated). Similarly, many methylation patterns may be compatible with X-chromosome inactivity. It is clear that several patterns of methylation at *HpaII* sites surrounding γ -globin sequences are associated with inactivity at these loci in clonal populations of fibroblasts (Fig. 3b, d).

The variety and instability of X DNA methylation patterns observed in normal human cells may explain our inability to reactivate loci on the inactive X chromosome with 5-azaC. Because 5-azaC did not induce extensive demethylation and because methylation of *HpaII* sites seems to occur spontaneously in these cells, we may not have achieved sufficient demethylation at critical sites for X-chromosome reactivation. It is significant that the human X chromosome in the foreign environment of the mouse-human hybrid cell is probably less methylated than in normal human cells, and that the pattern is

more stably transmitted. Therefore, it is not surprising that all 5-azaC-induced reactivation events have occurred in transformed mouse and chick cells^{16-18,26,27}, where the demethylation would be more extensive and more stable.

Even if methylation is involved in maintaining X-chromosome inactivity, it may be only one of several components influencing transcriptional activity. For example, the accessibility of DNA to methylases and the subsequent methylation at particular sites may depend on chromatin structure. The identical *HpaII*-generated banding pattern seen in male subclones suggests a stable higher order determinant of DNA methylation. Similarly, the more stable methylation pattern for the human X in interspecies hybrid cells may reflect changes in the chromatin structure attributable to a foreign environment or associated with transformation (for example, acquisition of immortality, heteroploidy). Chromatin differences of this kind might explain the extensive variability in amount and distribution of 5-methylcytosine between species⁵. Further evidence for more than one level of regulation underlying X-chromosome activity are recent observations that alleles on the inactive X in normal human cells, when derepressed, produce less enzyme than those on the active X³². The precise relationship between X inactivation and DNA methylation remains unknown.

We thank Joyce Axelman for assistance with the azacytidine studies and Joseph B. Wiley for assistance with purification and identification of recombinant phage. This work was supported by NIH grant HD 05465.

Received 16 September; accepted 17 December 1981.

- Holliday, R. & Pugh, J. E. *Science* **187**, 227-232 (1975).
- Sager, R. & Kitchin, R. *Science* **184**, 426-433 (1978).
- Riggs, A. D. *Cytogenet. Cell Genet.* **14**, 9-25 (1975).
- Lyon, Mary F. *Biol. Rev.* **47**, 1-35 (1972).
- Ehrlich, M. & Wang, R. Y. *Science* **212**, 1350-1357 (1981).
- Razin, A. & Riggs, A. D. *Science* **210**, 604-610 (1980).
- Liskay, R. M. & Evans, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4895-4898 (1980).
- Comings, D. C. *Expt Cell Res.* **74**, 383-390 (1978).
- McGhee, J. D. & Ginder, G. D. *Nature* **280**, 419-420 (1979).
- Mandel, J. L. & Chambon, P. *Nucleic Acids Res.* **7**, 2081-2103 (1979).
- van der Ploeg, L. H. T. & Flavell, R. A. *Cell* **19**, 947-958 (1980).
- Shen, C. K. J. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6634-6638 (1980).
- Pollock, Y., Stein, R., Ragin, A. & Cedar, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6463-6467 (1980).
- Wigler, M., Levy, D. & Perucho, M. *Cell* **24**, 33-40 (1981).
- Singer, J., Roberts-Ems, J., Luthardt, F. W. & Riggs, A. D. *Nucleic Acids Res.* **7**, 2369-2385 (1979).

- Groudine, M., Eisenman, R. & Weintraub, H. *Nature* **292**, 311-317 (1981).
- Compere, S. J. & Palmiter, R. D. *Cell* **25**, 233-240 (1981).
- Mohandas, T., Sparkes, R. S. & Shapiro, L. J. *Science* **211**, 393-396 (1980).
- Wolf, S. F., Maren, C. E. & Migeon, B. R. *Cell* **21**, 95-102 (1980).
- Waalwijk, C. & Flavell, R. A. *Nucleic Acids Res.* **12**, 4631-4641 (1978).
- Scott, A. F., Phillips, J. A. & Migeon, B. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4563-4565 (1979).
- Migeon, B. R., Brown, T., Axelman, J. & Migeon, C. I. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6339-6343 (1981).
- Benton, D. W. & Davis, R. W. *Science* **196**, 180-182 (1977).
- Lawn, R. M. et al. *Cell* **15**, 1157-1174 (1978).
- Migeon, B. R., Sprengle, J. A. & Do, T. T. *Cell* **18**, 637-641 (1979).
- Taylor, S. M. & Jones, P. A. *Cell* **17**, 771-779 (1979).
- Jones, P. A. & Taylor, S. M. *Cell* **20**, 85-93 (1980).
- Kahan, B. & DeMars, R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1510-1514 (1975).
- Clough, W. D. & Davidson, R. L. *Science* (in press).
- Wolf, S. F. & Schlessinger, D. *Biochemistry* **16**, 2783-2791 (1977).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Migeon, B. R., Wolf, S. F., Maren, C. E. & Axelman, J. (in preparation).
- Wilson, J. T. et al. *Nucleic Acids Res.* **5**, 563-581 (1978).

Vitamin A and pattern formation in the regenerating limb

M. Maden

Division of Developmental Biology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Vitamin A and its derivatives, the retinoids, cause specific alterations in the proximo-distal pattern of regenerating axolotl limbs. In some cases complete limbs grow from carpal level amputations. If retinoids are responsible for changing the positional information of limb cells, these compounds may represent valuable probes for investigating the molecular basis of pattern formation.

VARIOUS experimental approaches have been used to study pattern formation. The original approach, begun in the last century, is to alter the spatial relationships of the system under study by transplanting tissue to different sites and observing the resultant disruption of the pattern. In the limb, the system studied here, this approach has revealed organizing regions such as the zone of polarizing activity^{1,2}, the polarizing zone³ and the apical ectodermal ridge^{1,4}, and the beginnings of a theoretical framework into which the phenomena can be fitted⁵. An alternative approach is to search for substances which alter pattern formation in specific ways, for example, the effect of lithium on early development⁶. Unfortunately, in this case subsequent investigations did not prove particularly valuable in advancing our understanding of the molecular basis of pattern formation. Here I report that vitamin A and its derivatives, the retinoids, alter pattern formation specifically in the proximo-distal axis of regenerating axolotl limbs. If these effects are due to changes in the positional information of limb cells caused by retinoids, these compounds may be used in future research to elucidate the molecular nature of positional information.

Niazi and Saxena⁷ studied the effects of retinol palmitate on hindlimb bud regeneration of *Bufo* tadpoles. In addition to its inhibitory properties, they found several cases of duplicated or triplicated regenerates, some of which consisted of more than those distal parts of the limb removed by amputation. These results seemed not only to contradict the rule of distal transformation⁸, but also were unique in increasing the potency of blastemal cells as all other substances which have been administered to regenerating limbs have caused deficiencies^{9,10}. Thus I decided to investigate further this phenomenon in the regenerating limbs of larval axolotls, *Ambystoma mexicanum*, an animal commonly used for regeneration studies.

Effect of vitamin A

To confirm the existence of the effect described by Niazi and Saxena, I amputated the forelimbs of 14 animals through the mid-radius and ulna. The animals were immediately placed in a solution of retinol palmitate (15 IU ml⁻¹, made up in tap water) for 12 days and then transferred to normal water. An equal number of control animals were allowed to regenerate in pure tap water. After 1 month the limbs were fixed and stained. In contrast to the high mortality caused by hypervitaminosis A in *Bufo* tadpoles⁷ (and confirmed by my own results on *Rana* tadpoles), axolotls seemed immune. All experimental animals survived and seemed even healthier than normal.

In control limbs, all the elements removed by amputation were regenerated (Fig. 1a), whereas after retinol palmitate treatment, every limb (28 in all) was abnormal in the proximo-distal sequence of regenerated elements. The severity of the effect varied and could be separated into four classes (Table 1). The least severely affected were those limbs in which there was an extra row of carpals (Fig. 1b). In the next group an additional length of radius and ulna had been intercalated, which fused

with the stump radius and ulna resulting in an abnormally long segment (Fig. 1c). The third class comprised limbs in which a complete additional radius and ulna had been intercalated, resulting in a tandem repeat of that segment (Fig. 1d). The most severe effect was the additional presence of the distal part of the humerus (Fig. 1e). Most of the abnormalities observed (46%; Table 1, top row) consisted of extra carpals, the least severe effect.

In all other aspects, however, these experimental regenerates were normal: the digits were in the normal sequence with the

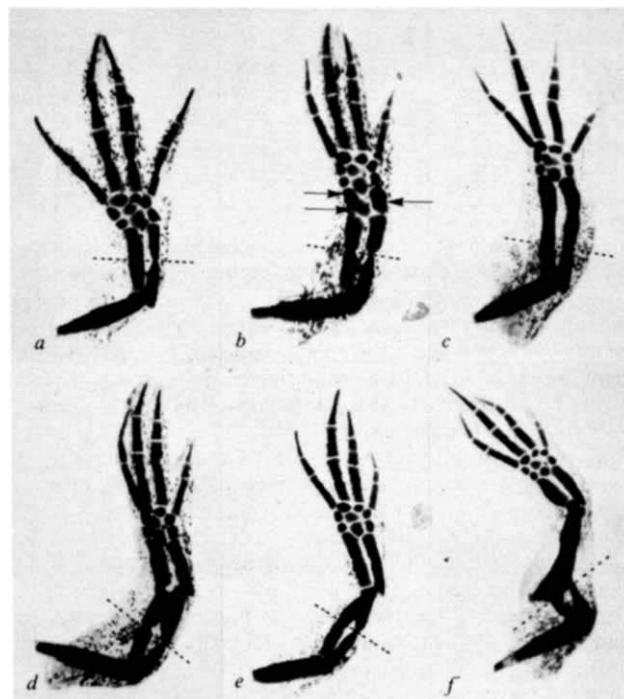


Fig. 1 Cleared whole mounts of regenerated limbs stained with Victoria blue. All limbs were amputated through the mid-radius and ulna, the broken line marking the amputation plane. *a*, Control. The elements removed by amputation (distal radius and ulna, 8 carpals and 4 digits) have been replaced. *b–e*, Limbs treated with 15 IU ml⁻¹ retinol palmitate for 12 days. *b*, Three extra cartilages (arrows) have been produced which resemble carpals. *c*, An abnormal length of radius and ulna has been regenerated making this segment grossly out of proportion. *d*, A complete extra radius and ulna has been produced in tandem. *e*, The radius and ulna have fused into the distal end of the humerus with a perfect elbow joint followed by a complete radius and ulna and hand. *f*, A limb regenerated after 15 days in 75 IU ml⁻¹ retinol palmitate; here a complete limb has been produced distal to the amputation plane. Magnifications: *a*, $\times 9$; *b*, $\times 8$; *c*, $\times 7.6$; *d*, $\times 7.6$; *e*, $\times 7$; *f*, $\times 6$.

Table 1 Effects of varying concentrations and lengths of exposure to vitamin A (retinol palmitate) after forelimb amputation through the mid-radius and ulna

Length of exposure (days)	Concentration of vitamin A (IU ml ⁻¹)	No. of cases	Type of regenerate					
			Normal (Fig. 1a)	Extra carpals (Fig. 1b)	Extra part radius and ulna (Fig. 1c)	Extra full radius and ulna (Fig. 1d)	Extra part humerus (Fig. 1e)	Extra full humerus (Fig. 1f)
12	15	28	0	13(46%)	7(25%)	5(18%)	3(11%)	0
10	30	10	0	0	7(70%)	3(30%)	0	0
15	30	8	0	0	1(13%)	2(25%)	6(62%)	0
10	75	10	0	0	0	5(50%)	5(50%)	0
15	75	8	0	0	0	0	1(13%)	7(87%)

Length of exposure is from the day of amputation. The types of regenerate are exemplified in Figs 1, 2 and described in more detail in the text.

correct phalangeal number, there were no alterations in the anteroposterior axis (compare Fig. 1a and b-e) nor any duplications or triplications as found in *Bufo* tadpoles⁷, and serial sections of four limbs revealed normal muscle patterns in the extra segments. Thus only the proximo-distal sequence of elements seemed to be affected in that varying lengths of extra pattern were intercalated between the amputation plane and the appropriate distal level.

Effect of increasing concentration and length of exposure

Four groups of five animals each had their forelimbs amputated as before, and were placed in 2× concentration (30 IU ml⁻¹) or 5× concentration (75 IU ml⁻¹) of retinol palmitate for 10 or 15 days. Neither concentration caused increased mortality and the effects of these treatments on pattern formation are shown in Table 1. Increasing the concentration caused a greater amount of pattern to be intercalated between the amputation plane and the appropriate distal level (compare rows 1, 2 and 4) such that after 10 days in 75 IU ml⁻¹, half the limbs had tandem repeats of the radius and ulna as well as the distal humerus. Similarly, increasing the time of exposure caused more pattern to be intercalated so that after 15 days in 75 IU ml⁻¹, a new class of abnormality appeared (Table 1, bottom row)—a complete limb regrew from the amputation plane, resulting in a proximo-distal sequence of elements of humerus, proximal radius and ulna (stump), full humerus, radius and ulna, carpals and digits (Fig. 1f).

Thus by increasing the concentration of retinol palmitate and the length of exposure, the level at which the limb was repeated became more and more proximal, finally resulting in a complete limb being regenerated from the distal amputation plane.

Morphological observations

While observing the animals during these experiments it became apparent that retinol palmitate was inhibiting the rate of regeneration. Normal regeneration in control animals progressed rapidly—cone stage on day 10, notch stage by day 12 and by day 18 four digits were present in outline. In the presence of retinol palmitate a progressive retardation of regeneration with increasing concentration was revealed and in 75 IU ml⁻¹, there were no external signs of regeneration. Immediately after removal to pure water, blastema formation commenced, consistent with the known inhibitory effects of retinoids on cell division. However, during the period in which there was no blastema development, preparations for the dramatic changes in pattern formation must, nevertheless, have been occurring.

Other amputation levels

The above experiments were repeated in animals whose forelimbs had been amputated either through the mid-humerus or through the carpals.

Amputation through the humerus (56 cases in all) gave the same results as those described above, that is, increasing the concentration of and length of exposure to retinol palmitate caused the level at which the limb was repeated to become more proximal. For example, after 12 days at the lowest concentration (15 IU ml⁻¹) the regenerates had abnormally long humeri,

whereas after 15 days in 30 or 75 IU ml⁻¹, 67% and 87% respectively of the cases regenerated a complete new limb from the amputation plane (Fig. 2a).

After amputation through the carpals (46 cases in all), many regenerates had digital deficiencies such as missing or fused digits at the lower concentrations. This was the only series of experiments in which these abnormalities occurred consistently. At 15 and 30 IU ml⁻¹ the only other effect on regeneration was the production of an extra row of carpals (for example, see Fig. 1b). However, at 75 IU ml⁻¹, six out of eight limbs regenerated complete limbs from the carpals (Fig. 2b). Furthermore, in two of these cases the most proximal element regenerated was not just the proximal epiphysis of the humerus, but also the glenoid cavity and part of the scapula (Fig. 2c).

By comparing the regenerates obtained from these three amputation levels of the forelimb, one further conclusion was drawn. The concentration of retinol palmitate needed to induce the regeneration of a complete limb decreased as the amputation plane became more proximal, that is, only the longest time at the highest concentration (75 IU ml⁻¹) led to regeneration of complete limbs from the carpal level, whereas 30 IU ml⁻¹ produced the same result from amputations at the humerus level.

Effect on hindlimbs

The whole range of experiments was repeated on a total of 133 hindlimbs. At high concentrations of retinol palmitate, complete limbs were regenerated from amputations through the mid-femur or mid-tibia and fibula (Fig. 3a) and with lower concentrations and shorter exposure times the extent of serial repetition decreased. In the cases where whole limbs regenerated, not only was there a complete femur, but also a

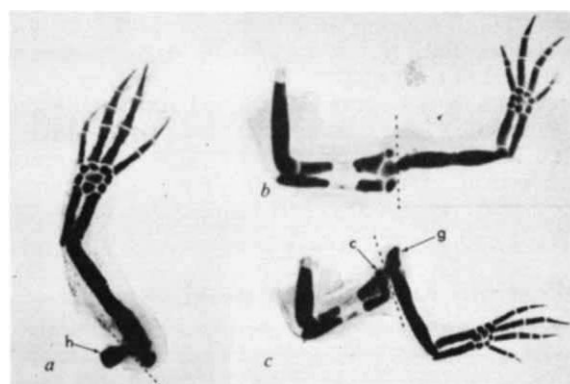


Fig. 2 Forelimbs placed in retinol palmitate (75 IU ml⁻¹) for 15 days. Broken lines mark the amputation planes. a, Proximal to the amputation plane is the remaining small part of the original humerus (h). Distally a complete limb has regenerated. b, Amputation through the carpals. Three carpals remain in the stump (radiale, intermedium and ulnare) and distally a complete limb has regenerated. c, Same as b but in this case part of the pectoral girdle (g) including the glenoid cavity has also been produced. Magnifications: a, ×6; b, ×5; c, ×5.

Table 2 Effects of 15-day exposure to various retinoids after amputation through the mid-radius and ulna or tibia and fibula

Compound		No. of cases	Type of regenerate						Relative effectiveness
			Normal	Extra carpals	Extra part radius and ulna	Extra full radius and ulna	Extra part humerus	Extra full humerus	
Retinol palmitate	(0.6 mM)	20	3(15%)	0	3(15%)	8(40%)	6(30%)	0	54
Retinol acetate	(0.6 mM)	13	0	0	12(92%)	0	1(8%)	0	43
Retinol palmitate	(0.12 mM)	20	15(75%)	0	5(25%)	0	0	0	10
Retinol acetate	(0.12 mM)	20	18(90%)	0	2(10%)	0	0	0	4
Retinol	(0.12 mM)	16	0	0	4(25%)	3(19%)	8(50%)	1(6%)	68
Retinoic acid	(0.12 mM)	15	0	0	0	3(20%)	9(60%)	3(20%)	80

Same classes of abnormalities as in Table 1. All compounds were obtained from Sigma. Retinol palmitate, retinol acetate and retinoic acid were dissolved in water, retinol in dimethyl sulphoxide.

socket, acetabulum and what appeared to be a small part of the pelvic girdle (Fig. 3a, b). Complete limbs were not regenerated from tarsal level amputations even at 75 IU ml⁻¹ as in the case of forelimbs. Subsequent experiments revealed that a concentration of 150 IU ml⁻¹ was needed for this to occur (Fig. 3b). This confirmed a trend which was apparent from a comparison of fore- and hindlimb experiments, that any one concentration and time had a greater effect (that is, caused regeneration to commence from a more proximal level) in the forelimb than in the hindlimb.

Effectiveness of other retinoids

I tested retinol palmitate, retinol acetate, retinol and retinoic acid to determine their comparative efficacy in inducing pattern abnormalities. Two concentrations were used: 0.6 mM (equivalent to 75 IU ml⁻¹ retinol palmitate) and 0.12 mM (equivalent to 15 IU ml⁻¹ retinol palmitate). Five animals whose fore- and hindlimbs had been amputated through the radius and ulna or tibia and fibula were immersed in each of the test solutions for 15 days (a total of 160 limbs). At the highest concentration, all the animals kept in retinol died, as did three out of five in retinoic acid with the remaining two not regenerating. One animal in retinol acetate died. Thus only retinol palmitate and retinol acetate could be compared in this experiment (Table 2). In 0.6 mM retinol palmitate, as before there was a spread of abnormalities, with most producing an extra radius and ulna in tandem (Fig. 1d) or an extra humerus also (Fig. 1e). The effects of retinol acetate were less severe, most consisting of an abnormally long radius and ulna (Fig. 1c).

At the lower concentration (0.12 mM; Table 2, rows 3–6), retinol palmitate had only minor effects; most of the regenerates were normal and the remainder had an abnormally long radius and ulna. Again retinol acetate was less effective. However, in 0.12 mM retinol the majority of limbs had an extra part humerus (Fig. 1e), and in retinoic acid, a further 20% of cases had complete limbs regenerating from the amputation plane even at this low concentration. Scoring the severity of pattern abnormalities from 1 (extra carpals) to 5 (extra full humerus) gave a measure of the relative efficacy of these four retinoids (Table 2, last column) and the following hierarchy was revealed: retinol acetate < retinol palmitate < retinol < retinoic acid. Retinoic acid was thus the most active of the four retinoids tested.

Implications for pattern formation

The most obvious conclusion from these results is that the regenerated limbs described above all contravene the 'rule of distal transformation'⁸. It has been shown that cells from any one level can only generate more distal limb levels during intercalary regeneration^{11,12} or regeneration from a reversed stump^{13,14} nor can they be forced to transform proximally in normal conditions¹². However, it is now clear that if conditions are changed by the addition of retinoids this rule no longer applies.

As the profound changes brought about by retinoids occur while blastemal development and presumably cell division is inhibited, it follows that changes in pattern do not necessarily

require cell division. They can be totally independent of growth, although subsequent proliferation is needed to manifest any alterations that have occurred. Thus pattern formation and growth control may be specified by different mechanisms. I have argued elsewhere¹⁵ that limb regeneration consists of two phases, an early morphallactic phase during which positional values are reassigned without cell division, followed by a period of growth when the remaining discrepancies are evened out. This scheme agrees in all respects with the results reported here.

One possible explanation of the results is that the time delay itself is responsible for the extra pattern generated. However, a similar retardation occurs after amputated limbs have been denervated, a procedure which removes the source of neurotrophic factor and inhibits cell division¹⁶, but in contrast to retinoid treatment, when nerves are allowed to regrow into such limbs, only those elements that were removed by amputation are regenerated¹⁷. Therefore time delays *per se* do not cause changes in pattern.

Possible mode of action of retinoids

Research into the cellular effects of retinoids has mainly involved mammalian tissues and even within this constraint many contradictory results have been reported. Thus extrapolation to amphibian tissues is of limited value, but perhaps two effects are noteworthy. One is the blocking of cell communication by gap junctions¹⁸ (although an increase in gap junction formation has also been noted¹⁹). Cellular communication must surely have a vital role in the normal process of

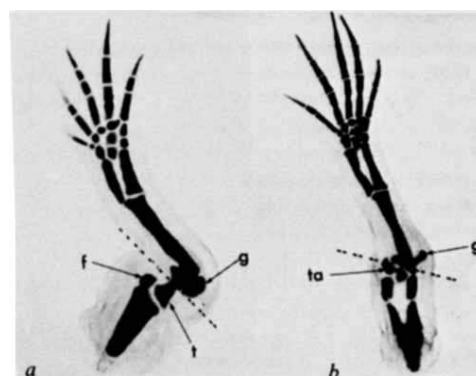


Fig. 3 Hindlimbs regenerated after vitamin A treatment. Broken lines mark the amputation planes. *a*, Retinol palmitate (75 IU ml⁻¹) for 15 days. The remaining parts of the tibia (*t*) and fibula (*f*) can be seen and not only has a whole limb been produced distally but also part of the pelvic girdle, the acetabulum (*g*). *b*, Amputation through the tarsals and retinol palmitate (150 IU ml⁻¹) for 15 days. The remaining tarsals (*ta*) can be seen and a complete new hindlimb including part of the girdle (*g*) has regenerated distally. Magnifications: *a*, ×6; *b*, ×6.

regeneration and any interference could cause pattern abnormalities. The other concerns speculations on the role of the cell surface in pattern formation. It has been suggested that positional information is encoded in the lengths of sugar chains on surface glycoproteins²⁰ and retinol does indeed act as a sugar donor in glycosyl transfer reactions²¹. In the nearest experimental system to that described here, that is, chick or mouse

limb bud mesenchyme, retinoids cause specific alterations in cell-surface glycoproteins^{22,23}. These speculations must await further research, but hopefully we now have a probe for investigating the molecular nature of positional information during pattern formation, at least in the limb.

I thank Katriye Mustafa for technical assistance in all aspects of this work.

Received 12 October 1981; accepted 18 January 1982.

1. Saunders, J. W. & Gasseling, M. T. in *Epithelial-Mesenchymal Interactions* (eds Fleischmajer, R. & Billingham, R. E.) 78–97 (Williams & Wilkins, Baltimore, 1968).
2. Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199–202 (1975).
3. Slack, J. M. W. *Nature* **261**, 44–46 (1976).
4. Stocum, D. L. *Differentiation* **3**, 167–182 (1975).
5. Wolpert, L. *J. theor. Biol.* **25**, 1–47 (1969).
6. Lallier, R. *Adv. Morph.* **3**, 147–196 (1964).
7. Niaz, I. A. & Saxena, S. *Folia Biol., Kraków* **26**, 1–8 (1978).
8. Rose, S. M. in *Regeneration 20th Growth Symp.* (ed. Rudnick, D.) 153–176 (Ronald, New York, 1962).
9. Dearlove, G. E., Lopez, S. & Dresden, M. H. *J. exp. Zool.* **194**, 413–428 (1975).
10. Bazzoli, A. S., Manson, J., Scott, W. J. & Wilson, J. G. *J. Embryol. exp. Morph.* **41**, 125–135 (1977).

11. Stocum, D. L. *Dev. Biol.* **45**, 112–136 (1975).
12. Maden, M. *J. Embryol. exp. Morph.* **56**, 201–209 (1980).
13. Butler, E. G. *J. Morph.* **96**, 265–281 (1955).
14. Wallace, H. *J. Embryol. exp. Morph.* **56**, 309–317 (1980).
15. Maden, M. *J. theor. Biol.* **69**, 735–753 (1977).
16. Singer, M. *Am. Zool.* **18**, 829–841 (1978).
17. Schotté, O. E. & Butler, E. G. *J. exp. Zool.* **97**, 95–121 (1944).
18. Pitts, J. D., Burk, R. R. & Murphy, J. P. *Cell Biol. int. Rep. Suppl. A* **5**, 45 (1981).
19. Elias, P. M., Grayson, S., Caldwell, T. M. & McNutt, N. S. *Lab. Invest.* **42**, 469–474 (1980).
20. Slack, J. M. W. *J. theor. Biol.* **82**, 105–140 (1980).
21. Lotan, R. *Biochim. biophys. Acta* **605**, 33–91 (1980).
22. Hassell, J. R., Pennypacker, J. P., Yamada, K. M. & Pratt, R. M. *Ann. N.Y. Acad. Sci.* **312**, 406–409 (1978).
23. Pennypacker, J. P., Lewis, C. A. & Hassell, J. R. *Archs Biochem. Biophys.* **186**, 351–358 (1978).

LETTERS TO NATURE

X-ray observations of the 1980 Cygnus X-1 'high state'

Y. Ogawara*, K. Mitsuda*, K. Masai†, J. V. Vallerga‡, L. R. Cominsky‡, J. M. Grunsfeld‡, J. S. Kruper‡ & G. R. Ricker‡

*Institute for Space and Astronautical Science, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan

†Department of Astrophysics, Nagoya University, Chikusa-ku, Nagoya, Japan

‡Department of Physics and Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Observations of intensity transitions in the X-ray emission from the black-hole candidate Cygnus X-1 (see ref. 1 and refs therein) were made by the Hakucho satellite² during the summer of 1980 and are reported here. In addition, we report the results of hard (20–120 keV) X-ray observations of Cyg X-1. These data were obtained during the 'high' intensity state by balloon-borne mercuric iodide detectors, and are compared with data obtained at similar energies approximately 1 month earlier (A. Sheepmaker *et al.*, in preparation) during the 'low' intensity state.

Cyg X-1 was observed to be in a 'high' state by the Uhuru satellite³ from December 1970 (its first observation of the source) to April 1971 and again during 1975 by the Ariel V (refs 4–6), and OSO-8 instruments⁷, as well as the Astronomical Netherlands Satellite⁸. Subsequent observations at many different X-ray energies have indicated that Cyg X-1 has been in an extended 'low' state since early 1976. In his review of Cyg X-1 (ref. 1), Oda has summarized the light curves and spectral information which were published up to 1977. During the transition to a 'high' state, the soft (< 10 keV) and hard (> 10 keV) X-ray intensities seem to be anticorrelated, that is, a 'high' state below 10 keV is associated with a substantial decrease in the X-ray intensity above 10 keV.

For the work reported here, soft X-ray observations (1–30 keV) were made with the scanning y-axis detectors on board the Hakucho satellite. The y-axis detector system consists of a pair of Xenon-filled proportional counters (SFX-VI and SFX-V2) with a $1.7^\circ \times 32^\circ$ FWHM field of view. Each of the two counters has an effective area of 32 cm². Only the SFX-V1 data were used in this analysis as the SFX-V2 field of view contained Cyg X-3 during most of the observations. The satellite spins at 6 r.p.m., so that 3,000 transits of a given source are available during the five daily 10-min passes above the tracking station at Uchinora, Japan. The source transits are superposed and corrected for

aspect transmission and Earth occultation. Background is also subtracted. (For further details on the Hakucho satellite see ref. 9.)

Figure 1 shows part of the X-ray light curve obtained with the Hakucho y-axis detectors during 1980. Each point represents the daily average X-ray intensity from Cyg X-1. Two 'high' states are clear: the first lasted at least 15 days, and was followed by a decrease over ~25 days to the pre-transition level; a second increase was observed ~4 months later. This transient event resulted in the highest measured intensity shown in Fig. 1 and began after JD 2444551.66 but before JD 244553.56 (binary phase 0.05 and 0.39 respectively). Thus the rise time was less than 1.9 days. This increase was quite brief, with the 'high' state persisting for only 2 days. The erratic variability seen during the 1980 observations is similar to that observed in 1975 and reported by Holt *et al.* using the Ariel V All Sky Monitor⁵.

Also indicated in Fig. 1 are the dates when hard X-ray (20–200 keV) balloon observations of Cyg X-1 took place. The 'low' state observations (8 May 1980) were performed by the MIT/Leiden Balloon Group and will be reported separately (A. Scheepmaker *et al.* in preparation). The 'high' state observations (28 June 1980) constituted the first use of a mercuric iodide X-ray detector¹⁷ in an astronomical experiment. The detector consisted of 11 HgI₂ crystals with a total effective area of 7.6 cm². The detector assembly (crystals and preamps) was enclosed within a cylindrical NaI scintillator well (active anticoincidence shield) which rejected charged particles and γ rays. The preflight energy resolution of the detectors ranged from 2–6 keV FWHM at 60 keV and remained stable during the 13-h flight from Palestine, Texas. The gondola containing the detector was

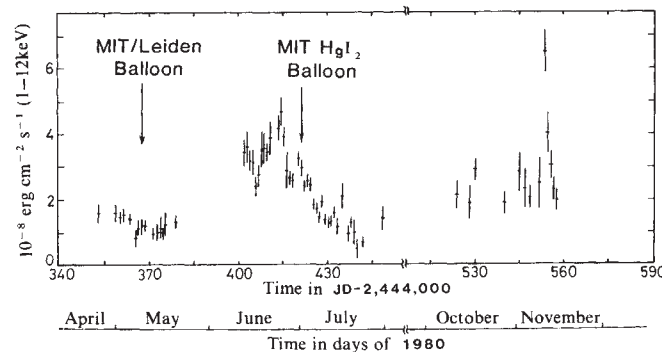


Fig. 1 Soft X-ray (1–12 keV) light curve of Cyg X-1 obtained with the Hakucho satellite. Two 'high' states are evident. The first began before 10 June 1980 and the second transient event occurred in November 1980. Also shown are the dates of simultaneous observations in hard X-rays by balloon-borne instruments.

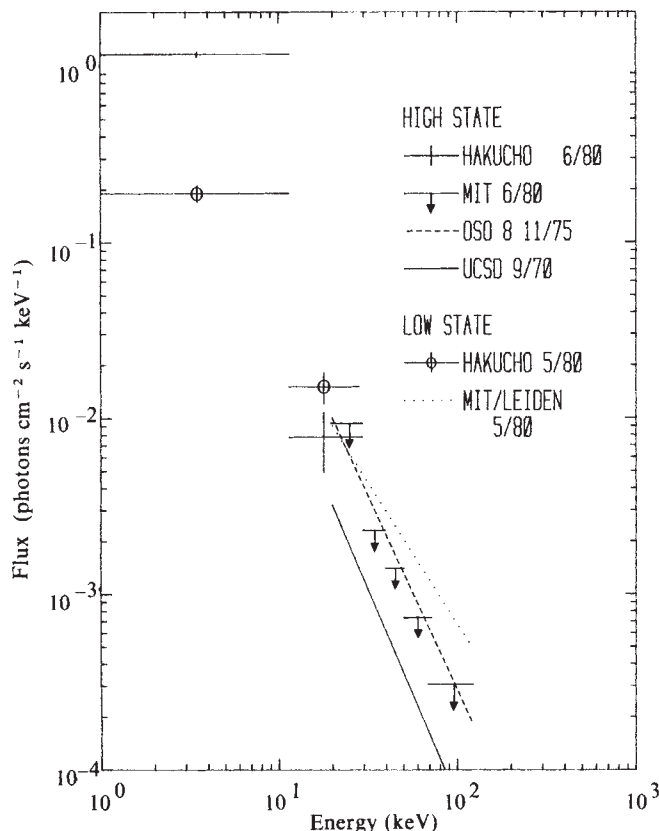


Fig. 2 X-ray spectra taken before and after the 'high' state transition of Cyg X-1 in late May/early June 1980. Data were obtained with the Hakucho satellite and with balloon-borne hard X-ray detectors. The Hakucho data points for high and low states are averages for 12–30 June 1980 and 1–19 May 1980, respectively. Also included are representative hard X-ray spectra taken during previous high states^{7,11}.

launched on an 11×10^6 ft³ balloon at 00.15 UT on 28 June 1980 and floated west for 13 h at an average altitude of 40 km (~ 2.9 g cm⁻²). Maximum exposure to Cyg X-1 occurred at 08.30 UT 28 June 1980. The detectors were calibrated using an on-board ²⁴¹Am source for 2 min out of every hour. The counts for each separate HgI₂ detector were binned in 64-s intervals in coarse energy channels, 10 keV wide. The detector counts were then summed to give the total counts per energy channel per 64 s.

Total exposure to the X-ray source was calculated as a function of time, taking into account aspect (azimuth, latitude, sidereal time, and field of view of individual HgI₂ detectors), atmosphere and detector window transmission, detector efficiency, and dead time. The maximum exposure to Cyg X-1 at transit was 37% over the 50–70 keV energy range. For each energy channel, an effective time on source, $t_{\text{eff}} = \sum_i t_i \epsilon_i$, was calculated and used to convert observed counts to a photon flux at the top of the atmosphere. Here, t_i refers to an individual time bin and ϵ_i is the total exposure, including atmospheric and aspect corrections as discussed above. There was less than a 2σ detection in all energy bands.

Cygnus X-1 passed through the field of view in 95 min resulting in 83 min of 'on source' data (that is, 5 min were occupied by calibrations and 7 min were lost due to telemetry dropouts); 40 min before and after this period were averaged and used as background.

Figure 2 shows the 2σ upper limits derived for the hard X-ray spectrum during the 28 June 1980 Cyg X-1 'high' state observation. Also indicated are the results from the 8 May 1980 hard X-ray observations by Scheepmaker *et al.* (in preparation) which were obtained when Cyg X-1 was in the 'low' state, as well as representative hard X-ray measurements during the 1971 and 1975 'high' states. The 1–12 keV and 12–30 keV X-ray intensities from the Hakucho observations have been averaged over

the high state and the low state separately. As can be seen in Fig. 2, the hard X-ray intensity is anticorrelated with the soft X-ray intensity; the pivot point occurs for an energy below 30 keV. The best fit power law for the Hakucho measurements gives a photon number spectral index of 1.5 ± 0.1 for the low state and 2.7 ± 0.2 for the high state. Using detectors on HEAO 3, Ling *et al.* (ref. 10 and personal communication) also recorded a continuous decrease in the hard X-ray intensity (48–183 keV) from 19 May 1980 until at least 2 June. The spectral anticorrelation reported here was also observed in previous low to high state transitions^{7,11} and seems to be a consistent feature of the bimodal behaviour of Cyg X-1.

Previous observations have established that the spectrum varies with time during the different intensity states and is not well represented by a simple relation (for example, power law or exponential). This has led to the idea that the soft component of the spectrum may originate in the outer optically thick region of an accretion disk surrounding the central compact object, whereas the hard component arises in an optically thin, hot inner region¹². The observed time scales for the transition to, and decay from, a 'high' state are consistent with theoretical calculations for the evolution of a standard accretion disk¹³, and suggest fluctuations in the accreted stellar wind or changes in the optical depth of the gas at the outer edge of the disk (see refs 1, 14 and refs therein). Spectral anticorrelation has also been used to explain the energy dependence of the short term variability (milliseconds to seconds)¹⁵. A numerical model proposed by Ichimaru¹⁶ shows that two physically distinct states can exist in the accretion disk depending on whether the radiative loss is smaller or larger than the viscous heating near the outer boundary (that is, a transition is expected to occur for $M/T^2 \sim 5 \times 10^3$ g s⁻¹ K⁻² where M is the accretion rate and T is the gas temperature at the outer edge of the disk). Thus, the middle part of the disk may become optically thick (high state) or optically thin (low state).

We conclude that Cygnus X-1 returned to its high state in 1980 after an apparent 4.5-yr interval in the low state. Note that this time period is similar to the 4-yr interval between the first and second observed transitions^{3,4}. We have established that the soft and hard X-ray intensities are anticorrelated in the transition between the two distinct states and we believe that any theoretical model of the bimodal behaviour of Cyg X-1 must account for this spectral anticorrelation.

We thank all the members of the Hakucho team for their contributions to the observations; Drs H. Inoue, K. Makishima and T. Murakami for help in the data analysis and preparation of the manuscript; Robert Kubara and the launch crew at the National Center for Atmospheric Research; and Dick Lynn, Carol Ortale, Wayne Schneppe and John Warren for assistance. A. Scheepmaker provided data from the MIT/Leiden 1980 balloon flight before publication. L.R.C. acknowledges support from Zonta International during part of this work. This research was sponsored in part by NASA under grants NAGW-14 and NSG7339.

Received 23 October; accepted 7 December 1981.

- Oda, M. *Space Sci. Rev.* **20**, 757–813 (1977).
- Oda, M. & Hakucho Team *IAU Circ. No.* 3491 (1980).
- Tananbaum, H. *et al. Astrophys. J. Lett.* **177**, L5–L10 (1972).
- Holt, S. S., Boldt, E. A., Kaluzienski & Serlemitsos, P. J. *Nature* **256**, 108–109 (1975).
- Sanford, P., Ives, J. C., Bell Burnell, S. J., Mason, K. O. & Murdin, P. *Nature* **256**, 109–111 (1975).
- Holt, S. S., Kaluzienski, Boldt, E. A. & Serlemitsos, P. J. *Nature* **261**, 213–215 (1976).
- Dolan, J. F., Crannell, L. J., Dennis, B. R., Frost, K. J. & Orwig, L. E. *Nature* **267**, 813–815 (1977).
- Heise, J. *et al. Nature* **256**, 107–108 (1975).
- Kondo, I. *et al. Space Sci. Instrum.* **5**, 211 (1981).
- Ling, J. C., Mahoney, W. A., Riegler, G. R., Wheaton, W. A. & Jacobson, A. S. *Bull. Am. astr. Soc.* **13**, 549 (1981).
- Matteson, J. L., Mushotzky, R. F., Paciesas, W. S. & Laros, J. G. in *X-Ray Binaries* (eds Boldt, E. & Kondo, Y.) 407–424 (NASA SP-389, 1975).
- Thorne, K. S. & Price, R. H. *Astrophys. J. Lett.* **195**, L101–L105 (1975).
- Lightman, A. P. *Astrophys. J.* **194**, 419–427 (1974).
- Eardley, D. M., Lightman, A. P., Shakura, N. I., Shapiro, S. L. & Sunyaev, R. A. *Comments Astrophys.* **7**(5), 151–160 (1978).
- Doi, K. *Nature* **287**, 210–212 (1980).
- Ichimaru, S. *Astrophys. J.* **214**, 840–855 (1977).
- Vallerga, J. V., Ricker, G. R., Schneppe, W. & Ortale, C. *IEEE Trans. Nucl. Sci.* **NS-29** (1982).

Quantum conformal fluctuations in a singular space-time

T. Padmanabhan & J. V. Narlikar

Tata Institute of Fundamental Research, Bombay 400 005, India

The cosmological solutions of Einstein's general relativistic equations lead inevitably to space-time singularities¹. However, general relativity is only an approximation to a fully quantized theory of gravity and we need to consider whether singularity persists in the quantum domain. Although a full quantum theory of gravity has not yet been developed, we show here that the above question can be tackled in a simplified model where only the conformal degree of freedom is quantized. Previous applications of this technique had shown that in specific cases the quantum conformal fluctuations (QCF) from the classical solutions diverge at the classical singularity, thus rendering the classical solution physically meaningless^{2,3}. Recently one of us (J.V.N. ref. 4) has generalized this result to cover all dust cosmologies. Here we show that this conclusion is applicable to even more general types of cosmological singularities.

To arrive at this result we draw on the mathematical details of previous work; in particular, we make extensive use of the formalism developed in ref. 4.

Classical equations of general relativity are derived from the variation of the action S given by

$$S = \frac{1}{16\pi} \int R \sqrt{-g} d^4x + \int \mathcal{L}_m \sqrt{-g} d^4x \quad (1)$$

We have taken $c = 1$ and $G = 1$. As we are discussing quantum phenomena we will also take $\hbar = 1$, thus making the Planck length the basic unit. $\mathcal{L}_m$ is the lagrangian density for matter.

In the thin sandwich problem⁵ of classical gravity, if data are given on initial and final hypersurfaces Σ_1 and Σ_2 , $\delta S = 0$ determine the geometry of the space-time sandwiched between Σ_1 and Σ_2 . The initial and final data are specified by 3-geometries $^{(3)}\mathcal{G}_1$ and $^{(3)}\mathcal{G}_2$. In the quantum problem there is no unique sequence of 3-geometries leading from Σ_1 to Σ_2 . Instead we have a probability amplitude $K[^{(3)}\mathcal{G}_2, \Sigma_2; ^{(3)}\mathcal{G}_1, \Sigma_1]$ for the system to have the specified initial and final data. The kernel K is formally written as the Feynman functional integral

$$K[^{(3)}\mathcal{G}_2, \Sigma_2; ^{(3)}\mathcal{G}_1, \Sigma_1] = \int \exp iS[\mathcal{G}] \mathcal{D}\mathcal{G} \quad (2)$$

where $S[\mathcal{G}]$ is the action for an arbitrary geometry $\mathcal{G}$ which satisfies the given boundary conditions. If we are considering only the conformal degrees of freedom we have to take all geometries given by the metrics

$$g_{ik} = \Omega^2 \bar{g}_{ik} \quad (3)$$

where Ω is a function of space and time ($i = 0, 1, 2, 3$), while $\bar{g}_{ik}$ is the metric tensor describing the solution of the classical problem (all physical quantities pertaining to the classical solution carry an overbar).

Using a time coordinate t to label the spacelike hypersurfaces Σ_1, Σ_2 as t_1, t_2 we may rewrite equation (2) as

$$K[\Omega_2, t_2; \Omega_1, t_1] = \int \exp iS[\Omega] \mathcal{D}\Omega \quad (4)$$

with

$$S[\Omega] = \frac{1}{16\pi} \int (\bar{R}\Omega^2 - 6\Omega_i\Omega^i) \sqrt{-\bar{g}} d^4x + \int \mathcal{L}_m \sqrt{-\bar{g}} d^4x \quad (5)$$

The lack of a unique space-time geometry in the quantum regime may also be expressed through a wavefunction Ψ which 'propagates' with the kernel K :

$$\psi(\Omega_2, t_2) = \int K[\Omega_2, t_2; \Omega_1, t_1] \psi(\Omega_1, t_1) \mathcal{D}\Omega_1 \quad (6)$$

A physical interpretation can be given to the above relation by

choosing a Gaussian wavepacket form for $\psi(\Omega_1, t_1)$ and then studying its dispersion as the final hypersurface Σ_2 approaches the singularity. For the examples discussed in refs 2 and 3, it was shown that while the 'mean' of the wavepacket stays on the classical solution $\Omega = 1$, its dispersion diverges as the classical singularity is approached.

It was shown in ref. 4 that the above results were in fact part of a general result which was valid for all singular space-times containing only dust. By a similar analysis it can be shown that whenever $S[\Omega]$ is a quadratic functional in Ω and its derivatives, the kernel can be written as

$$K[\Omega_2, t_2; \Omega_1, t_1] = F(t_2, t_1) \exp[i\bar{S}] \quad (7)$$

where $F(t_2, t_1)$ depends only on t_2 and t_1 , and

$$\bar{S} = \sum_{P=1}^2 \sum_{Q=1}^2 \frac{3}{8\pi} \iint A_{PQ}(\mathbf{x}_P, t_P; \mathbf{x}'_Q, t'_Q) \times \Omega_P(\mathbf{x}_P) \Omega_Q(\mathbf{x}'_Q) d^3\mathbf{x}_P d^3\mathbf{x}'_Q \quad (8)$$

The coefficients A_{11} and A_{22} are given by certain integrals involving the retarded Green's function $G(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1)$ of a scalar wave equation determined by the background classical geometry and the source. In the dust case discussed in ref. 4, the scalar wave operator is simply $\square + \bar{R}/6$.

The interesting part of $\bar{S}$ in equation (8) is, however, the cross term containing A_{12} , which retains the memory of the initial wavefunction at $t = t_1$. The coefficient A_{12} is given by

$$A_{21}(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) = A_{12}(\mathbf{x}, t_1; \mathbf{x}_2, t_2) = [G(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1)]^{-1} \quad (9)$$

Suppose now that G diverges as the space-time singularity is approached. We then have $A_{12} \rightarrow 0$ and a decoupling of the final wavefunction from the initial one. This is precisely the state of complete uncertainty of the final state represented in the wavepacket case by infinite dispersion.

In ref. 4 the divergence of G at the singularity was deduced from the conformal invariance of the operator $\square + \bar{R}/6$. However, the result also seems to be valid for other cases. For example, it was shown recently⁶ that the QCF diverge for a Robertson-Walker universe containing a massless scalar field. The QCF also diverge at the singularity of the empty Kasner universe⁷. These results suggest that we should look for a proof of the divergence of G in a fairly general scenario as follows.

Belinskii *et al.*⁸ argued that a general approach to the space-time singularity can be described by a succession of Kasner type solutions, in which the Universe contracts along two axes and expands along the third. (We have reversed the direction of time so that $t \leq 0$ and the final singular state is described by $t = 0$.) Without going into detail we will simply accept the claim that number, 8, of arbitrary functions in the Belinskii *et al.*⁸ solution makes it a general solution of Einstein's equations.

Belinskii *et al.*⁸ have argued that near the space-time singularity the contributions from $\mathcal{L}_m$ can be neglected and the wave operator for our Green's function then becomes simply $\square$. We therefore have to show that in the space-time of their model the Green's function of the wave equation

$$\square \bar{G}(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) = [-\bar{g}(\mathbf{x}_1, t_1)]^{-1/2} \delta_4(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) \quad (10)$$

diverges as $t_2 \rightarrow 0$.

The line element of the Belinskii *et al.* model is given by

$$ds^2 = dt^2 - \gamma_{\alpha\beta} dx^\alpha dx^\beta, \quad \alpha, \beta = 1, 2, 3 \quad (11)$$

$$\gamma_{\alpha\beta} = \sum_{\mu=1}^3 |t|^{2p_\mu} l_{\mu\alpha} l_{\mu\beta} \quad (12)$$

where $p_\mu, l_{\mu\alpha}$ are functions of space coordinates. As in the Kasner universe, the p_μ satisfy

$$\sum_{\mu=1}^3 p_\mu = \sum_{\mu=1}^3 p_\mu^2 = 1 \quad (13)$$

Simple algebra shows that close to the singularity the Green's function is given by

$$G(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) \equiv \frac{\delta_3(\mathbf{x}_2, \mathbf{x}_1)}{f(\mathbf{x}_2)} \ln \left| \frac{t_1}{t_2} \right| \quad 0 > t_2 > t_1 \quad (14)$$

while $G=0$ otherwise. The function f depends on the space coordinates. We are interested in the behaviour of the Green's function as $t_2 \rightarrow 0$, the singular epoch.

From equation (14) we have $G \rightarrow \infty$ logarithmically as $t_2 \rightarrow 0$, leading to divergence of QCF. Matter terms are unimportant in the final stages of the Belinskii *et al.* solution so our conclusion should hold for general distributions of matter and energy.

Although we have quantized only the conformal degree of freedom, the fact that the QCF diverge at the classical singularity is sufficient to demonstrate that quantum uncertainty becomes so large near the classical singular epoch that the classical singularity ceases to have any significance. In analogy with the example of the hydrogen atom can we therefore expect explicit examples of quantum behaviour in the very early Universe? Recent investigations^{9,10} have shown that closed Friedmann models have stationary states and that the Planck length sets the lower limit to the 'size' of the Universe. Whether this is a general feature of quantum cosmology remains to be seen.

Received 2 November 1981; accepted 6 January 1982.

1. Hawking, S. W. & Ellis, G. F. R. *The Large Scale Structure of Space-time* (Cambridge University Press, 1973).
2. Narlikar, J. V. *Nature* **269**, 129 (1977).
3. Narlikar, J. V. *G. R. G. J.* **10**, 883 (1979).
4. Narlikar, J. V. *Foundn. Phys.* **11**, 473 (1981).
5. Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Freeman, San Francisco, 1973).
6. Padmanabhan, T. *Phys. Rev. D* (in the press).
7. Padmanabhan, T. Preprint (Tata Institute of Fundamental Research, Bombay, 1981).
8. Belinskii, V. A., Khalatnikov, I. M. & Lifshitz, E. M. *Adv. Phys.* **19**, 525 (1970).
9. Padmanabhan, T. & Narlikar, J. V. *Phys. Letts.* **A84**, 361 (1981).
10. Padmanabhan, T. Preprint (Tata Institute of Fundamental Research, Bombay, 1981).

Individual electrons detected after the interaction of ionizing radiation with gases

O. H. W. Siegmund, J. L. Culhane, I. M. Mason & P. W. Sanford

Mullard Space Science Laboratory, University College London, Dorking, Surrey RH5 6NT, UK

The interaction of ionizing radiation (X rays or high-energy particles) with gases results in the creation of electron-ion pairs. For low-energy X rays a small cluster of primary ionization is produced close to the point of the initial photoionization. High-energy particles, however, lose a small fraction of their energy in each of a number of ionizing collisions resulting in a long track composed of clusters of ionization. We report here that the light signals emitted by a parallel-plate proportional counter are fast enough to allow the individual electrons in the ionization clusters to be resolved. By counting the number of electrons for low energy X ray events we demonstrate an improvement in energy resolution of more than a factor of two over conventional proportional counters.

The mean number of electrons, $\bar{n}$, created by the photoelectric absorption of X rays in a gas is a function of the X-ray energy, E , and the mean energy required to create an ion pair, W , such that $\bar{n} = E/W$ (where $W \sim 26$ eV in Ar). In conventional gas proportional counters the electrons produced are accelerated by a strong electric field such that a Townsend charge avalanche results. At a constant mean charge gain, $\bar{A}$, the charge signal from the proportional counter is directly proportional to the X-ray energy. The charge signal pulse amplitude distribution which is observed results from the variation in n and in A . Fano¹ pointed out that the standard deviation in n is $\sigma_n = (F\bar{n})^{1/2}$, where F (the Fano factor) is ~ 0.17 for Ar. The total fractional deviation in the charge pulse amplitude, $\sigma_N/\bar{N}$, is given by,

$$\frac{\sigma_N}{\bar{N}} = \left[\left(\frac{\sigma_n}{\bar{n}} \right)^2 + \frac{1}{\bar{n}} \left(\frac{\sigma_A}{\bar{A}} \right)^2 \right]^{1/2} \quad (1)$$

where $\bar{N}$ is the mean number of final electron-ion pairs, and

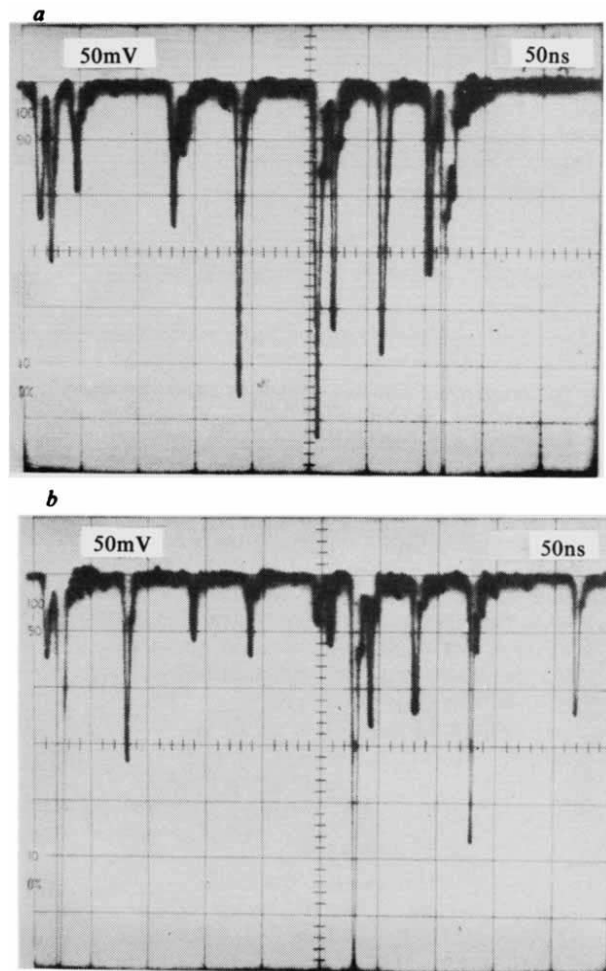


Fig. 1 Traces of the light pulses for two C-K X-ray events. The individual peaks correspond to avalanches initiated by single electrons. The gas used was Ar-CH₄(5%)-CO₂(5%)-N₂(7.5%), with a drift region field strength of ~ 10 V cm⁻¹ and a gas gain of 10^6 .

$\sigma_A/\bar{A}$ is the fractional deviation in the charge gain. We may write $f = (\sigma_A/\bar{A})^2$ for the variance in A (f is typically $0.6-0.7$ when $A = 10^4-10^5$). Rewriting equation (1) in the form $\sigma_N/\bar{N} = [(F+f)W/E]^{1/2}$, it is clear that the elimination of f would result in a factor of more than two improvement in energy resolution. This is accomplished in the gas scintillation proportional counter^{3,4}. The same result may be achieved by counting the individual initial electrons.

In the approach described here the light pulses emitted by Townsend avalanches in a parallel plate proportional counter⁵ are detected by a fast photomultiplier tube (PMT, EMI 9818 QBM, ~ 2 ns rise time). The PMT views the 1 mm-deep avalanche region through an optically transparent anode deposited on a fibre optic substrate^{5,6}. The gas mixture used was Ar-CH₄(5%)-CO₂(5%)-N₂(7.5%) at a pressure of 760 torr. Ar and N₂ in the mixture are responsible for the light emission. Ar emits in the range 7,000-8,000 Å and N₂ emits in the range 3,000-4,000 Å (from the second positive band of N₂), with an optimum light emission attained for of the order of 10% N₂ (ref. 6). CH₄ and CO₂ are used in the gas mixture to inhibit secondary electron emission processes, thus enabling high charge gains to be achieved, and to reduce the lateral diffusion of the electron cluster so that acceptable position resolution may be obtained in applications requiring imaging^{6,7}. Addition of CO₂ also reduces the electron drift velocity⁸ (A. J. F. Den Boggende and C. J. Schrijver, personal communication). Fairly low concentrations (5%) of CH₄ and CO₂ were used as these gases also suppress the light emission⁶.

For an X-ray event the overall duration of the light signal is set by the time taken for the initial electron cluster to cross the grid

which separates the drift and avalanche regions of the counter. Because each electron initiates an avalanche the overall light signal is composed of a number of light pulses (n). The ability to resolve the light pulses attributable to individual electrons depends on the number of electrons, the overall duration of the light signal and the width of the individual light pulses. As the electron velocities in the avalanche region are high the excitation time, and hence the rise time, of the individual light pulses is fast (~ 1 ns). It has also been found that the light pulse decay time is fast (~ 1 ns) because CO_2 and CH_4 heavily quench the excited states responsible for light emission by collisional de-excitation⁷. This time scale compares with a resolution of ~ 20 ns obtained by Walenta⁸ with the charge signal from a time expansion drift chamber. Walenta⁸ was able to resolve the ionization clusters for high energy charged particle events, but was not able to resolve individual electrons for 5.9-keV X-ray events, although this might have been possible if low energy (< 1 keV) X rays and very low electron drift velocities had been used. Light pulses from a counter of similar geometry to that used in this work were studied by Charpak *et al.*⁹ and show some structure for ^{55}Fe events, but the detector gas and operating conditions were not optimized for high time resolution.

In this study the overall duration of the light signal, required to be able to resolve the light pulses initiated individual electrons, was established by using a low field strength (10 V cm^{-1}) in the drift region of the counter, giving a low electron drift velocity ($\sim 1.5 \times 10^5 \text{ cm s}^{-1}$)⁷. However, this field strength is substantially less than that required for optimum position resolution ($> 100 \text{ V cm}^{-1}$)⁷ and results in an increase of a factor of ~ 3 in the lateral size of the electron cluster.

Figure 1 demonstrates the resolution of C-K X rays into several single electron events (the signals were displayed on a Tektronix 1 GHz oscilloscope). It is estimated (from the data in ref. 2) that for the gas mixture used $F \sim 0.17$ and $W \sim 26.5 \text{ eV}$ so the theoretical number of initial electrons is 10.6 ± 1.3 . This compares with an experimental value of 11 ± 1.6 , deduced by counting the number of light pulses which comprise each individual C-K X-ray event. For C-K X rays this method results in an energy resolution of $\sim 34\%$ FWHM, compared with a value of $\sim 85\%$ FWHM for typical proportional counters. For B-K (0.18 keV) X rays we obtain a value of 7 ± 1.5 electrons compared with a theoretical value of 7 ± 1.1 . This gives an experimental energy resolution of $\sim 50\%$ FWHM at B-K compared with $\sim 105\%$ FWHM for conventional counters. The energy resolution for the electron counting technique is

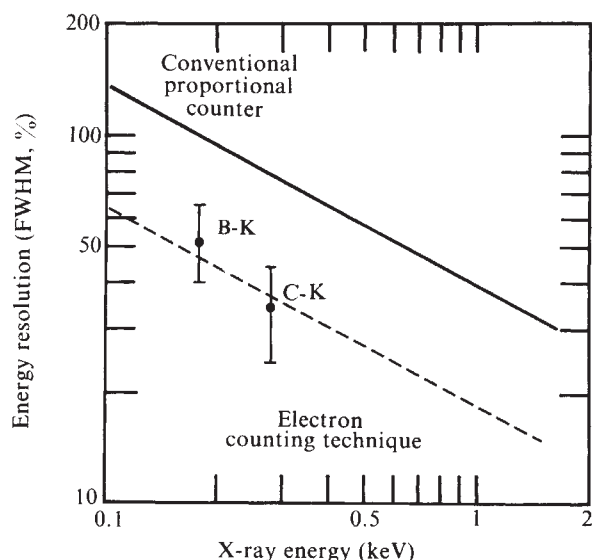


Fig. 2 Comparison of the FWHM energy resolution as a function of X-ray energy for a proportional counter with the predicted energy resolution for counting individual initial electrons, derived from measurements with C-K and B-K X rays.

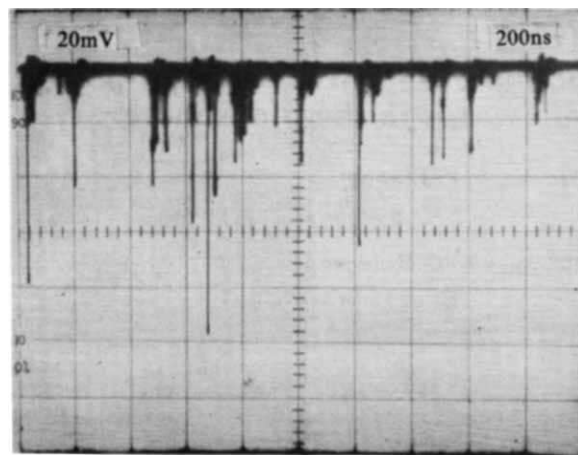


Fig. 3 Trace of the light pulse emitted for a ^{60}Co event showing the resolution into clumps of ionization, which are made up of a number of peaks corresponding to individual electrons.

compared with the energy resolution of a conventional proportional counter in Fig. 2.

^{60}Co ($\sim 1.2 \text{ MeV}$) γ rays were also observed with the detector. γ rays interact mainly with the walls, resulting in the ejection of high-energy electrons. These electrons are mostly within a factor of two of being minimum ionizing. The ionization tracks produced extend across the 3.5-mm deep drift region of the counter, therefore the overall light signals lasts about four times longer than for X-ray events, which produce small electron clouds. Figure 3 shows the structure of a typical ^{60}Co event in which ionization clusters may be discerned, each composed of several electrons.

In X-ray detection the increased light signal duration for high energy particle events allows identification and rejection of the latter. This ability has already been demonstrated with the integrated light signal^{5,6}. Background rejection may be accomplished more efficiently, however, by setting upper limits to the number of single electron avalanches and the duration of the light pulse train.

The practical realization of electron avalanche counting will require suitable pulse counting circuits. Although the theoretical limiting resolution, as determined by the Fano factor, is $\sim 37\%$ (B-K) and $\sim 30\%$ (C-K) it is anticipated that counting errors due to closely spaced single electron events will degrade the resolution achievable in practice. Results obtained with fast logic devices indicate that electronic counting can be at least as efficient as the technique used here. Therefore it seems possible that improvements in energy resolution of at least a factor of two over current proportional counters may be attained.

This improvement in energy resolution for a parallel plate proportional counter should find applications in low energy X-ray astronomy, where both high energy resolution and high spatial resolution are desirable⁶. The ability to count individual electrons, as demonstrated here, makes it possible to determine Fano factors for gases, as was suggested by Walenta⁸. The increased time resolution may also enable improved ionization cluster counting efficiency to be achieved for applications in high-energy charged particle physics.

We thank Professor R. L. F. Boyd for his encouragement. This work was supported by the SERC.

Received 4 November 1981; accepted 11 January 1982.

1. Fano, U. *Phys. Rev.* **72**, 26–29 (1947).
2. Alkhazov, G. D., Komar, A. P. & Vorob'ev, A. A. *Nucl. Instrum. Meth.* **48**, 1–12 (1967).
3. Alice, M., Alves, F., Policarpo, A. J. P. L. & Salete, M. *IEEE. Trans. nucl. Sci.* **NS-22**, 107–115 (1975).
4. Inoue, H., Koyama, K., Matsuoka, M., Tanaka, Y. & Tsunemi, H. *Nucl. Instrum. Meth.* **157**, 295–300 (1978).
5. Siegmund, O. H. W., Sanford, P. W., Mason, I. M., Culhane, J. L. & Cockshott, R. *Nature* **287**, 808–810 (1980).
6. Siegmund, O. H. W. *et al.* *IEEE. Trans. nucl. Sci.* **NS-28**, 478–482 (1981).
7. Siegmund, O. H. W. thesis, Univ. London (1981).
8. Walenta, A. H. *IEEE. Trans. nucl. Sci.* **NS-26**, 73–80 (1979).
9. Charpak, G., Majewski, S. & Sauli, F. *Nucl. Instrum. Meth.* **126**, 381–389 (1975).

Nonlinear reflectivity of high-power radio waves in the ionosphere

H. Kopka & P. Stubbe

Max-Planck-Institute für Aeronomie, 3411 Katlenburg-Lindau 3, FRG

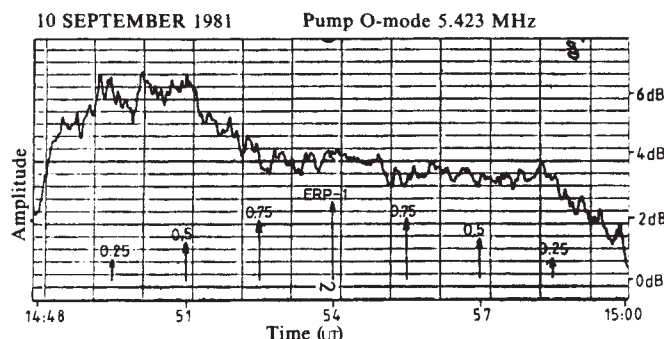
T. B. Jones & T. Robinson

Department of Physics, University of Leicester,
Leicester LE1 7RH, UK

An important result of earlier ionospheric modification experiments performed at Platteville, Colorado, was the discovery that the reflectivity of the ionosphere was greatly reduced in the reflection region of a high-power radio (pump) wave¹. Because the reflectivity of low-power diagnostic waves passing through the modified region with frequencies from just below the pump wave frequency up to the F region critical frequency were reduced, the phenomenon was termed wide-band absorption (WBA). Subsequent experimental and theoretical work (for review see ref. 2) has led to the conclusion that WBA is due to the scattering of radio waves from field-aligned irregularities into plasma waves. These irregularities are generated by the interaction of the pump wave and the ionospheric plasma but the mechanism by which this occurs is not completely understood. It is evident, however, that in the process the pump wave itself undergoes WBA. In a recent campaign using the Max-Planck Institute ionospheric modification facility at Ramfjordmoen, near Tromsø, Norway, the existence of WBA at high latitudes was confirmed by experiments using low-power diagnostic waves passing through the reflection region of the high-power radio wave^{3,4}. As we report here, during an experiment to measure the effect of WBA on the pump wave itself, the essentially nonlinear nature of the interaction between the high-power radio waves and the ionosphere was demonstrated in a new and striking way.

The experiment consisted of measuring the amplitude of the reflected 5.423-MHz pump wave at a receiver about 40 km south of the transmitter while the effective radiated power (ERP) of the pump wave was being changed. The changes in ERP consisted of 40 steps of 6.5-MW increases from 0 to 260 MW (full power) and then 40 incremental decreases of 6.5 MW down to 0 MW, the steps being made at 9-s intervals. One cycle was thus completed within 12 min and was repeated several times.

Figure 1 shows a record of the reflected pump signal during this experiment. The reflected signal increases suddenly during the first step from 0 to 6.5 MW ERP and then continues to increase but less rapidly, during successive increases in ERP. This behaviour is a result of the usual WBA effect and persists up to ~40–50% of full power. However, further increases in power produce a completely new type of behaviour in which the reflected signal decreases with increasing ERP.



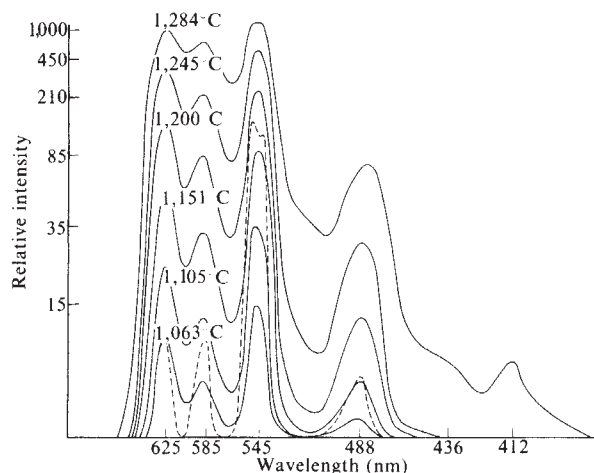


Fig. 1 Spectra obtained on cooling of $\text{Tb}(\text{PO}_3)_3$. ---, UV-excited spectrum of $\text{Tb}(\text{PO}_3)_3$.

the melt was allowed to cool freely in air in its crucible, a cooling rate of $\leq 2^\circ\text{C s}^{-1}$, no emission was observed. On casting from an initial melt temperature of $1,400^\circ\text{C}$, the centre temperature of the cast piece showed a cooling rate of $\sim 10^\circ\text{C s}^{-1}$ and the emission for a 100 g melt was visible for ~ 20 s. In another set of experiments small solid glass samples were heated to up to $1,100^\circ\text{C}$ and then quenched. If the samples were allowed to cool in air the initial cooling rate was $\sim 100^\circ\text{C s}^{-1}$ but no CIL was observed. However, if the samples were taken from the heat source and immediately placed in water, giving an initial cooling rate of $\sim 500^\circ\text{C s}^{-1}$, CIL was observed for samples having an initial temperature of 850°C or higher. No CIL was observed in cases where the initial temperature was $< 850^\circ\text{C}$. These results indicate that a certain quench rate is needed to observe CIL and that this quench rate is related to the initial temperature of the glass.

Several melts were made by melting a mixture of Tb_4O_7 and $\text{NH}_4\text{H}_2\text{PO}_4$ to determine the composition range over which the CIL effect could be observed. Additions, if any, were in the form of oxides, except in the case of sodium where the carbonate was used. The melts were carried out at $1,400^\circ\text{C}$ in Al_2O_3 crucibles after an initial preheating to drive off most of the NH_3 and H_2O and preheat the crucible before placing it in the melt furnace. The melts were generally cast after 1 h at the melting temperature.

In binary terbium phosphates, CIL was observed over a composition range of 16–25 mol % Tb_2O_3 . No melts were made with > 25 mol % Tb_2O_3 . Various oxides were added to the $\text{Tb}(\text{PO}_3)_3$ composition, at 5 mol %, to determine their effect on the CIL. Individual additions of MgO , Na_2O , and Al_2O_3 did not seriously affect the emissions. However, additions of oxides with variable valency states, such as Sb_2O_3 or CeO_2 , completely quenched the CIL effect, although UV-excited fluorescence was observed.

As an alternative to introducing the P_2O_5 as $\text{NH}_4\text{H}_2\text{PO}_4$, other melts were prepared from material formed by mixing H_3PO_4 and Tb nitrate solution. The terbium nitrate was made by digesting terbium oxide in concentrated nitric acid. The material was dried at $\sim 100^\circ\text{C}$ before placing it in the melt crucible. Melting was carried out in an Al_2O_3 crucible at $1,400^\circ\text{C}$ for 1 h, as with the other melts, and resulted in a clear fluid melt which cooled to a clear glass. Although it showed only a very faint

indication of CIL on pouring, UV-excited fluorescence was observed.

The intensity of the CIL in these melts could be increased greatly, to the level of the original glasses, by bubbling NH_3 or CO gas through the melts at $1,400^\circ\text{C}$. The CIL could be quenched again by bubbling O_2 gas through the melt.

During melting of the various compositions, it was noted that holding time at temperature also affected the CIL. The Na_2O -containing glass, which originally exhibited CIL, did not show CIL after an additional 3 h of melting. The Al_2O_3 -containing glass also showed a significant decrease in intensity of the CIL in similar conditions, as did $\text{Tb}(\text{PO}_3)_3$, particularly at long melting times, up to 22 h. The ability of oxygen to quench the CIL indicates that the decrease in intensity may be due to the slow oxidation of the melt from prolonged heating in air.

CIL was also photographically recorded for a europium metaphosphate melt, though the luminescence is in the red portion of the spectrum and therefore difficult to distinguish visually from the black body radiation.

Our results identify the emissions involved in the CIL as the normal Tb^{3+} fluorescence transitions, and provide some indication of the nature of the excitation process. One obvious possibility is excitation by the thermal black-body radiation. However, if the mechanism was a simple excitation by the UV portion of the continuum, a steady-state emission would be expected while holding the glass at elevated temperatures, and none is observed. Another possibility is a stress-induced triboluminescence. However, the CIL occurs at temperatures where the melt is still quite fluid and no stress could be developed.

The need for quenching to observe the CIL, and the dependency of high quench rates on starting temperature, indicate that the emission is the result of the decay of a temperature-dependent excited state. The fact that the cooling rate necessary to observe the CIL depends on starting temperature may indicate that the decay process is always radiative but during slow cooling the emission cannot be observed because of the relative magnitude of the background. The other possibility is a competition between radiative and nonradiative decay with the radiative process being favoured in rapid cooling conditions and the nonradiative process in slow cooling conditions.

One possible temperature-dependent mechanism is a valency change of the terbium. Only terbium oxide is known to show a large change in $\text{Tb}^{4+}/\text{Tb}^{3+}$ ratio as a function of temperature¹, and many other multivalent ions, such as iron² and cerium³, show similar changes even when incorporated in a glassy matrix. That oxidation of the melt, by batch additions (such as Sb_2O_3 , CeO_2) or by bubbling O_2 gas, quenches the CIL and reduction of the melt, by CO_2 or NH_3 gas, can revitalize the CIL indicates the sensitivity of the effect to the overall redox balance of the

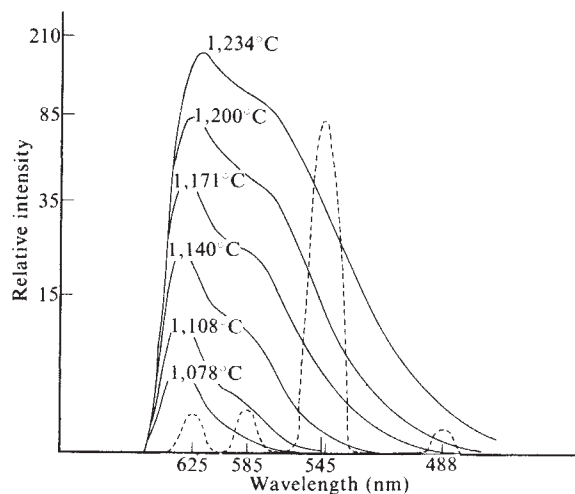


Fig. 2 Spectra obtained on cooling of $\text{La}(\text{PO}_3)_3$. ---, UV-excited spectrum of $\text{Tb}(\text{PO}_3)_3$.

Table 1 Effect of thermal history on CIL in $\text{Tb}(\text{PO}_3)_3$

Initial temperature ($^\circ\text{C}$)	Cooling rate ($^\circ\text{C s}^{-1}$)	
	CIL observed	CIL not observed
1,400	10	2.5
1,100–850	500	100
≤ 800	—	500

system. Therefore a temperature-dependent valency change may indeed contribute to the CIL effect. However, the intensity of the Tb^{3+} fluorescence in UV irradiation was essentially the same, for terbium phosphate glasses, regardless of whether CIL was observed or not, implying that the Tb^{3+} content of the glasses, at room temperature, was not greatly different, thus indicating that the redox changes are more complicated than a simple shift in $\text{Tb}^{4+}/\text{Tb}^{3+}$ ratio. It also is not clear from our results whether the quenching of the CIL by oxidation prevents formation of the excited state or shifts the decay process to a non-radiative path.

In conclusion, a unique form of luminescence was observed which was quite sensitive to the overall oxidation reduction balance of the melt and the rate of cooling of the melt. The experimental evidence indicates that the excitation mechanism requires species of a particular valence and may be due to a thermally induced shift in the oxidation reduction balance in the melt which results in the observed terbium luminescence. This may involve the terbium directly or may involve the phosphorus with the energy released being transferred to the terbium to produce the fluorescence.

Received 10 September; accepted 15 December 1981.

1. Lowe, A. T. & Eyring, L. *J. Solid State Chem.* **14**, 383–384 (1975).
2. Johnston, W. D. *J. Am. ceram. Soc.* **47**(4), 198–201 (1964).
3. Gottardi, V. *et al.* in *Advances in Glass Technology*, Proc. 6th int. Glass Congress, Washington, DC (Plenum, New York, 1962).

Slope anomaly in the vapour pressure curve of Hg

S. R. Hubbard

Lowestoft College of Further Education, St Peters Street,
Lowestoft NR32 2NB, UK

R. G. Ross

School of Mathematics and Physics, University of East Anglia,
Norwich NR4 7TJ, UK

The investigation of normally-metallic fluids at reduced densities has created considerable experimental and theoretical interest. Hensel¹ has presented evidence for transitions, related to electronic properties, in both the liquid and vapour phases of Hg, and predicted that these would affect the liquid-vapour transition. We present here evidence that the vapour pressure curve of Hg, $p(T)$, has an unusual and probably unique form in that the $\ln p$ against $1/T$ plot showed a sharp change of slope at temperature $T = T' = 1,361$ K and saturated liquid density $\rho_L = 10.7 \text{ g cm}^{-3}$. The latter is about the density at which Hensel¹ deduced that the electronic mean free path equalled the mean interatomic spacing, and which he took to be associated with the electronic-property related transition in the liquid phase.

Our experiments were made using an internally-heated pressure vessel pressurized with argon gas. The sample of Hg was contained in a closed cell, fabricated largely of Mo, which entered the internal furnace of the pressure vessel and which was also connected by means of high-pressure capillary tubing to a strain gauge type of pressure transducer outside of the pressure vessel. This transducer was used to measure the vapour pressure. The argon pressure in the pressure vessel was adjusted so as to be approximately equal to the measured vapour pressure of the Hg to prevent deformation of the cell under stress. The pressure transducer was calibrated using a dead-weight tester, and we estimate the inaccuracy in p as ± 0.4 MPa. The temperature was measured using several Pt/Pt13Rh thermocouples in close contact with the hot part of the cell where the liquid-vapour meniscus was located. Users of internally-heated pressure vessels are familiar with system-

atic errors in temperature measurement arising from convection in the pressure-transmitting medium, here argon gas. To correct for such error, we adjusted our temperatures so that, at 929 K, our data agreed with the relatively accurate, low-temperature vapour pressure data of Sugawara *et al.*². The required adjustment of temperature varied from 0 to 20 K in the various runs which were made.

We fitted our data to several different equations using a least-squares procedure. The simplest form of equation was $\ln p = A/T + B$. To obtain a good fit to our data, we needed different values of A and B above and below a temperature of 1,361 K where we observed a change of slope (Fig. 1). At low temperatures, our data in the range 742–1,271 K yielded

$$\ln(p/\text{MPa}) = -7,148(\text{K}/T) + 9.064 \quad (1)$$

Using the same procedure, the data of Sugawara *et al.*² in the temperature range 602–929 K yielded

$$\ln(p/\text{MPa}) = -7,132(\text{K}/T) + 9.038 \quad (2)$$

From the agreement of the corresponding parameters in equations (1) and (2), in particular that corresponding to the slope, we inferred that systematic error in temperature measurement in our experiments was constant, essentially independent of temperature, in any given run. Over the temperature range 742–929 K, where there was data overlap between our experiments and those of Sugawara *et al.*², pressures calculated using equations (1) and (2) differed at most by 0.88%.

The same fitting procedure and form of equation when applied to our data in the temperature range 1,389–1,719 K yielded

$$\ln(p/\text{MPa}) = -8,180(\text{K}/T) + 9.822 \quad (3)$$

To confirm the validity of our data at the higher temperatures, we used the most recent and reliable values of critical temperature, T_c , and critical pressure, p_c , obtained independently

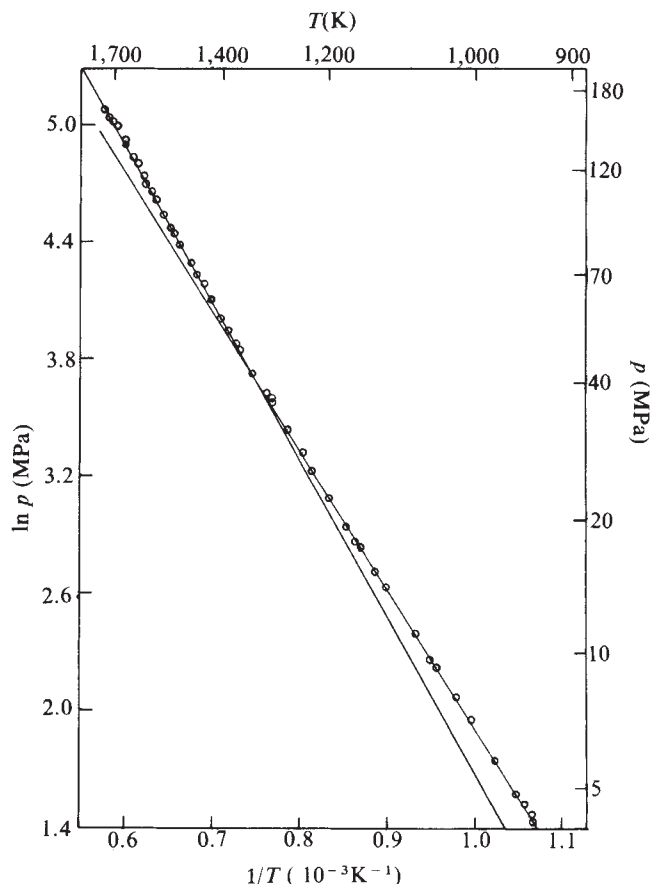


Fig. 1 Vapour pressure of Hg.

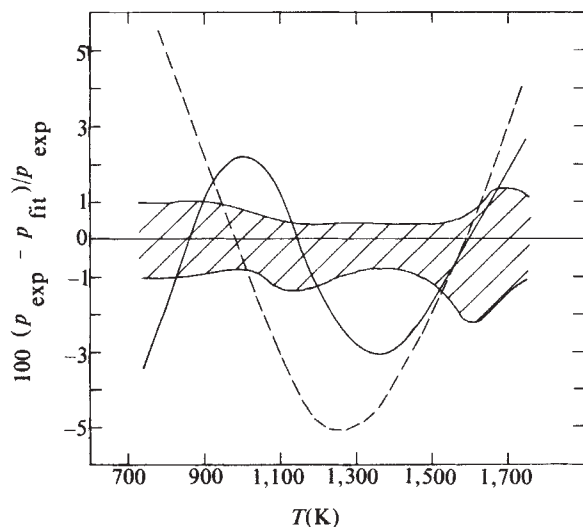


Fig. 2 Deviation between experimental and fitted values for vapour pressure of Hg: solid line, $\ln p = A/T + B + C \ln T$; broken line, $\ln p = A/T + B$; cross-hatched region, deviations scattered within this region both for Frost-Kalkwarf⁶ equation $\ln p = A/T + B + C \ln T + Dp/T^2$, and for equations (1) and (3) of the present work.

by other workers. Experiments by Neale and Cusack³ and Schönherr and Hensel⁴ yielded an experimental value $p_c = 170$ MPa, obtained by averaging 167.5 and 172 MPa. Substitution in equation (3) of values of T_c obtained by these workers yielded an average predicted value $p_c = 179$ MPa, that is, 9 MPa higher. Alternatively, substitution of the same values of T_c in equation (1) yielded an average predicted value $p_c = 151$ MPa, that is, 19 MPa lower than the average experimental value obtained independently. From such comparisons we inferred: (1) that our predicted value of p_c from equation (3) was in reasonable agreement with other work, (2) that our values of vapour pressure near p_c were not subject to serious unsuspected systematic error and (3) that linear extrapolation of low temperature data in a $\ln p$ against $1/T$ plot (equation (1)) yielded an improbably low predicted value of p_c . As we have found reasonable agreement of our data with that of other workers^{3,4} at $T = T_c$, and good agreement of the slope with the data of Sugawara *et al.*² at low temperatures, it follows that the $\ln p$ against $1/T$ plot must bend upward at intermediate temperatures.

The typical results shown in Fig. 1 were reproducible both on heating and cooling and for a range of different total Hg masses in the cell. The temperature at which the change of slope occurred (T') we estimated as $1,361 \pm 30$ K.

We also attempted to fit other forms of equation to our data; the results are shown in Fig. 2. A single equation of form $\ln p = A/T + B$, or the Kirchhoff equation $\ln p = A/T + B + C \ln T$, evidently yielded poor fits. The latter has been used successfully⁵ to fit high-temperature vapour pressure data for Cs, which showed only a slight curvature in a $\ln p$ against $1/T$ plot, concave towards the $1/T$ axis. The four-parameter implicit equation, $\ln p = A/T + B + C \ln T + Dp/T^2$, due to Frost and Kalkwarf⁶, gave as good a fit as two equations of form $\ln p = A/T + B$, and thus these two possibilities cannot be distinguished on numerical grounds.

Whatever the form of equation(s) used, the data shown in Fig. 1 clearly indicate a rapid change of slope at a temperature $\sim 1,361$ K. As already suggested by Hensel¹, liquid-vapour equilibrium would be expected to be affected by an electronic 'transition' in the liquid phase or vapour phase, or both. By using our value for T' together with well-established co-existence curve data⁷, we obtained a value $\rho_L = 10.7 \pm 0.3$ g cm⁻³ for the corresponding liquid phase density. This is about the density which has been taken^{1,8} to be that by which the electron mean free path has been reduced to about the mean interatomic

spacing, and which marks the beginning of a rapid reduction in the density of states at the Fermi level ($N(E_F)$) for lower densities. The latter densities have been referred to as corresponding to the strong scattering region. As the density of states should affect the thermodynamic properties, we conclude that the reduction of $N(E_F)$ with volumetric expansion of the liquid phase in the strong scattering region is the most likely cause of the different slope of the vapour pressure curve which we have found for temperatures greater than 1,361 K.

S.R.H. acknowledges financial support from the Suffolk County Council. We thank F. Hensel for a preprint.

Received 14 October 1981; accepted 18 January 1982.

1. Hensel, F. in *Proc. 8th Symp. on Thermophysical Properties*, Gaithersburg, (in the press).
2. Sugawara, S., Sato, T. & Minamiyama, T. *Bull. Jap. Soc. mech. Engr* **5**, 711-718 (1962).
3. Neale, F. E. & Cusack, N. E. *J. Phys. F: Metal Phys.* **9**, 85-94 (1979).
4. Schönherr, G. & Hensel, F. *Ber. Bunsenges. Phys. Chem.* **85**, 361-367 (1981).
5. Ewing, C. T. *et al. J. Chem. Engng Data* **15**, 508-510 (1970).
6. Frost, A. A. & Kalkwarf, D. R. *J. chem. Phys.* **21**, 264-267 (1953).
7. Ross, R. G. & Greenwood, D. A. *Prog. Mater. Sci.* **14**, 175-242 (1969).
8. Cusack, N. E. in *The Metal Non-metal Transition in Disordered Systems* (eds Friedman, L. R. & Tunstall, D. P.) 455-490 (Scottish Universities Summer School in Physics, Edinburgh, 1978).

Factors controlling the acidity of natural rainwater

R. J. Charlson* & H. Rodhe

Department of Meteorology, University of Stockholm, Arrhenius Laboratory, S-106 91 Stockholm, Sweden

It is often assumed¹⁻⁴ that the pH of natural rainwater is controlled by the dissociation of dissolved CO₂, has a value of 5.6, and that decreases below this are due to the addition of acidic components by human activity. However, decreases could be due to the removal by rainwater of naturally occurring acids from the air (notably H₂SO₄ in the natural portion of the sulphur cycle). Consideration of the cycling of water and sulphate through the atmosphere and the amount and composition of sulphate aerosol expected to be scavenged by a given amount of cloud water in remote locations indicates that, in the absence of basic materials (such as NH₃ and CaCO₃), average pH values of ~ 5 are expected to occur in pristine locations. This value must vary considerably due to variability in scavenging efficiencies as well as geographical patchiness of the sulphur, nitrogen and water cycles. Thus, pH values might range from 4.5 to 5.6 due to variability of the sulphur cycle alone. Because of widespread concern regarding the acidification of rain, it is important to understand the factors controlling the pH and composition of natural rainwater. We suggest here that there are integral constraints imposed by (1) the requirement for mass continuity in each elemental cycle and (2) the relative concentrations of soluble species and liquid water in cloudy air; these factors must be mutually consistent.

Although the NH₃/NH₄⁺ part of the nitrogen cycle is important (NH₃ being the dominant atmospheric gaseous base) it is not yet possible to consider it quantitatively. Similarly, the NO/NO₂/HNO₃ cycle is inadequately measured. However, HNO₃ will always add to the acidity of rainwater due to H₂SO₄. We can make a preliminary estimate of the concentrations of strong acids in rain in remote locations by first studying the sulphur cycle alone. The maximum or potential acidity thus estimated neglects neutralization by bases such as NH₃ or CaCO₃ which will increase pH. However, as both NH₃ and basic dust concentrations are often low (such as in maritime air) this approach has actual application.

Estimates of the natural flux of sulphur through the atmosphere have been based on indirect arguments: for example,

*Present address: Environmental Engineering and Science Program, Department of Civil Engineering, University of Washington, Seattle, Washington 98195, USA.

using river run-off data⁵, and more direct measurements of deposition by precipitation⁶, of emissions of sulphur-containing gases from the ocean surface⁷, coastal sediments⁸, or volcanoes⁹. The global total of such fluxes—disregarding anthropogenic emissions and those associated with the formation of sea spray—is still uncertain but recent estimates^{6,10} are in the range $35\text{--}110 \times 10^{12} \text{ g S yr}^{-1}$, which is comparable with the present anthropogenic emissions¹¹ of $\sim 100 \times 10^{12} \text{ g yr}^{-1}$. For illustrative purposes we may assume a value of $60 \times 10^{12} \text{ g yr}^{-1}$ for the global emission of natural, non-sea spray sulphur, but it is uncertain by at least a factor of two in either direction. If this sulphur were evenly distributed over the globe and one-half of it were deposited as sulphate in precipitation¹², the resulting *pH* would be ~ 5.2 (assuming the remaining sulphur is present as H_2SO_4 in the rainwater).

Note that the 'global sulphur cycle' is far from global. Because of the limited residence time in the atmosphere of the major sulphur-bearing compounds (H_2S , $(\text{CH}_3)_2\text{S}$, SO_2 and SO_4^{2-})¹² and an uneven geographical and temporal distribution of sources and sinks, the sulphur cycle will exhibit large variations in space and time. Examples include volcanic emissions of SO_2 ; production of $(\text{CH}_3)_2\text{S}$ in the surface water of the oceans and its subsequent release to the atmosphere that is likely to be coupled to the biological productivity of the ocean water and therefore vary with it; and the possibility of enhanced sulphur emissions in coastal areas (see, for example, ref. 13). (OCS, while important in terms of concentration, is assumed to be less important in this context due to a smaller flux.) Consequently, the global averages need not be very representative for any particular location. As a first step towards examining the factors controlling this variability, the cycle can be evaluated over spatial scales compatible with the average residence time in the atmosphere of the important sulphur compounds, that is, a few hundred to a few thousand kilometres¹⁴. Studies of regional sulphur budgets have already been attempted for regions with very large industrial emissions^{15,16}.

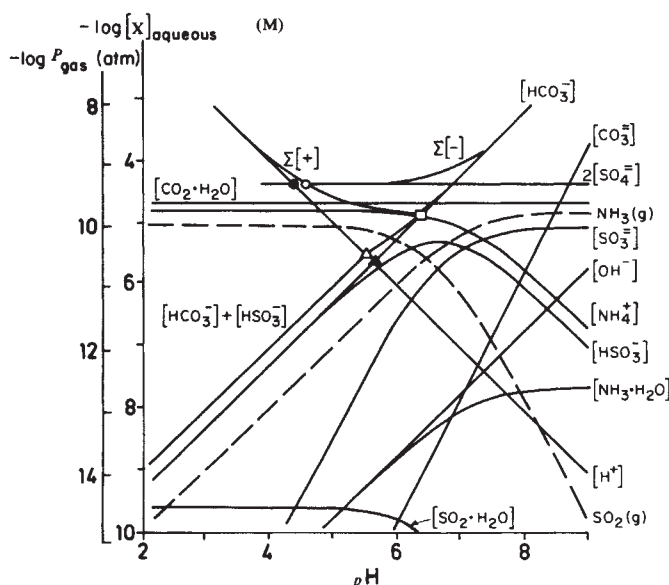


Fig. 1 Master variable diagram for the system $\text{H}_2\text{O}(\text{l})$, CO_2 , SO_2 , NH_3 and SO_4^{2-} . These input conditions represent a possible cloud in a remote, unpolluted area. Assumptions: $T = 5^\circ\text{C}$; $L = 0.5 \text{ g m}^{-3}$ liquid cloud water; $\text{CO}_2 = 340 \text{ p.p.m.v.}$; $\text{SO}_2 = 100 \text{ p.p.t.v.}$; $(\text{NH}_3 + \text{NH}_4^+) = 0.13 \text{ } \mu\text{g m}^{-3}$ as NH_3 ; $(\text{SO}_4^{2-})_{\text{aerosol}} = 1.0 \text{ } \mu\text{g m}^{-3}$, all dissolved in cloud. Charge balances: $\circ$, pH 4.6 for whole system; $\bullet$, pH 4.4 in absence of NH_3 (and NH_4^+); $\square$, pH 6.2 in absence of SO_4^{2-} ; $\triangle$, pH 5.5 in absence of both NH_3 and SO_4^{2-} ; $\blacktriangle$, pH 5.6 in absence of NH_3 , SO_4^{2-} and SO_2 . $[X]$ denotes aqueous concentration (M). $\text{SO}_2(\text{g})$ and $\text{NH}_3(\text{g})$ represent the gas phase (---). Equations and equilibrium constants have been taken from refs 17, 35.

Table 1 Potential acidity (*pH*) of atmospheric liquid water: effects of dissolved aerosol SO_4^{2-} and liquid water content of cloud

SO_4^{2-} ($\mu\text{g m}^{-3}$)	L (g m^{-3})		
	0.1	0.5	2.5
0.04	5.1	5.4	5.5
0.2	4.4	5.1	5.4
1.0	3.7	4.4	5.1

Concentrations of CO_2 and SO_2 are assumed to be constant and equal to 340 p.p.m.v. and 100 p.p.t.v. respectively. Total $\text{NH}_3 = 0$.

We conclude that substantial variations around the global average value of the potential acidity attributable to naturally circulating sulphur should be expected. The absolute magnitude of such variations is hard to estimate without a better understanding of the characteristics of the regional sulphur cycles, but variation by a factor of five either way seems possible. The corresponding *pH* values would then lie in the range 4.5–5.6. If the natural part of the atmospheric sulphur cycle is not spatially uniform by even a factor of two, then the *pH* of rainwater unaffected by human activity could be < 5.0 in some regions. (Note that the concentration of hydrogen ion in rainwater will be influenced by other compounds as well; for example, the $\text{NH}_3/\text{NH}_4^+$ cycle may be as variable as the sulphur cycle, and probably not in phase with it.)

Another approach to estimating the composition of rainwater is to utilize existing data on aerosol composition and liquid water content of clouds. Although the degree to which the composition thus calculated may reflect that of rain itself is not known precisely, the composition of rainwater, cloud water and the aerosol and trace gases that went into the cloud must be related.

Equilibria exist in clouds among several trace gases, their respective dissolved species and liquid water, and there are further inputs of aerosols containing sulphates, soil-derived dust and sea salt to cloud droplets. The sulphate aerosols are of particular interest because they often occur as H_2SO_4 partly neutralized by NH_3 (ref. 17). An approximate empirical formula of NH_4HSO_4 exists in remote, maritime submicrometre aerosols¹⁸ and, as such particles can be incorporated into cloud droplets, sulphate aerosols are potentially important sources of strong acid in rain. If we neglect the soil dust and sea salt, given the 3 Henry's law expressions for NH_3 , SO_2 and CO_2 , 6 equilibria for dissociation of these and water, 3 mass conservation equations for NH_3 and NH_4^+ , $\text{S}(\text{IV})$ and carbonate species and one equation of charge balance, a system of 13 equations can be solved for the equilibrium conditions¹⁷. Figure 1 is a Sillén or master variable diagram¹⁹ for the system NH_3 , SO_2 , CO_2 , H_2SO_4 and $\text{H}_2\text{O}(\text{l})$ in a pristine cloud, in which the concentration of each dissolved species and the partial pressure of the trace gases are given as functions of *pH*. When the 9 equilibrium statements are combined with the conservation equations, 12 separate curves result, portraying the concentration of each of the species. In addition, $[\text{SO}_4^{2-}]$ is given as *pH* independent at $\text{pH} > 3$, while $\text{CO}_2(\text{g})$ is omitted and the sum of $[\text{HSO}_3^-] + [\text{HCO}_3^-]$ is included. The charge balance then is given at the *pH* where the sums of positive and of negative charges are equal ($\Sigma[+] = \Sigma[-]$). For this set of conditions (see Fig. 1 legend), at $\text{pH} = 4.6$ there is charge neutrality (open circle in Fig. 1) and sulphate is the dominant anion; $[\text{NH}_4^+]$ is slightly $< [\text{H}^+]$, $[\text{HCO}_3^-] \approx [\text{HSO}_3^-] \approx 10^{-6.8}$ and $\ll [\text{SO}_4^{2-}]$. Also, the partial pressure scale is set so that the curves for $\text{SO}_2(\text{g})$ and $\text{NH}_3(\text{g})$ have the same limiting ordinate value as $[\text{SO}_3^{2-}]$ and $[\text{NH}_4^+]$ respectively. Thus the mass continuity expressions for these two can also be followed graphically as functions of *pH*, showing that if $\text{pH} < 5.5$, most $\text{S}(\text{IV})$ is $\text{SO}_2(\text{g})$ while NH_4^+ is the dominant ammonia species.

This plot can also be used to show what equilibrium *pH* would occur in the absence of various components. Only in the absence of NH_3 , SO_2 and SO_4^{2-} does *pH* equal 5.6; higher values occur with the absence of SO_4^{2-} alone, while *pH* decreases if NH_3 is removed. Increasing CO_2 slightly decreases the *pH* in those

cases where SO_4^{2-} is low. (Doubling CO_2 causes at most a decrease in pH of 0.15.) Decreasing SO_4^{2-} or increasing NH_4^+ by even small amounts increases the pH significantly. Complete oxidation of all the SO_2 in the cloud would raise $[\text{SO}_4^{2-}]$ to $2.8 \times 10^{-5} \text{ M}$ and decrease the equilibrium pH to 4.4, with all other variables constant. Thus, although these input conditions were selected as possible in a remote, unpolluted area, they are clearly variables to which pH is very sensitive.

To study further the relationship between the chemical compositions of aerosols, cloud water and rainwater and to estimate the effect of their variability on pH , consider the simplest case of in-cloud scavenging of a pre-existing sulphate aerosol with no in-cloud SO_2 oxidation. The cloud water composition is then given by²⁰

$$[\text{SO}_4^{2-}]_{\text{cloud}} = \frac{\epsilon (\text{SO}_4^{2-})_{\text{aerosol}}}{96L} \quad (1)$$

where $[\text{SO}_4^{2-}]_{\text{cloud}}$ is the molar sulphate concentration in cloud water, 96 is the molecular mass of SO_4^{2-} , $(\text{SO}_4^{2-})_{\text{aerosol}}$ is the mass concentration of sulphate aerosol in air (g m^{-3}), ϵ is the scavenging efficiency for sulphate aerosol particles in the cloud and L is the liquid water content (l m^{-3}) of the cloud. For soluble particles of 0.1–1.0 μm radius, ϵ is expected to be²⁰ 0.5–1.0.

Allowing for dilution (D) and evaporation (E) of the cloud and rainwater, the sulphate concentration in rainwater becomes:

$$[\text{SO}_4^{2-}]_{\text{rain}} = (\text{SO}_4^{2-})_{\text{aerosol}} \frac{\epsilon DE}{96L} \quad (2)$$

Data from the OECD study on long-range transport of air pollutants²¹ indicate a value of $\epsilon DE/L$ in the range $(0.6\text{--}1.6) \times 10^3 \text{ m}^3 \text{ kg}^{-1}$. If ϵ , D and E were unity, this would correspond to a liquid water content of $\sim 1 \text{ g m}^{-3}$, a value which is reasonable compared with observations of the liquid water content of precipitating clouds²² and consistent with other published values of washout ratios^{23,24}.

Table 1 lists values of expected pH of rain as a function of $(\text{SO}_4^{2-})_{\text{aerosol}}$ and L for $\epsilon \times D \times E = 1$. The effects on pH of compounds other than CO_2 and a background concentration of SO_2 of 100 p.p.t.v. (parts per 10^{12} by vol.)²⁵ are disregarded. For simplicity, SO_2 is assumed not to react. The L values used embrace most measured values of the liquid water content of clouds²². Generally, the lower values may be more representative of fogs and stratiform clouds²⁶ and the higher values of convective clouds, especially the rain-producing ones. Data from remote locations on the concentration of aerosol sulphate (excluding the contribution from sea spray) are sparse but values vary from $\sim 0.05 \mu\text{g SO}_4^{2-} \text{ m}^{-3}$ in remote parts of South America²⁷ to of the order of $1 \mu\text{g SO}_4^{2-} \text{ m}^{-3}$ in the marine boundary layer^{28,29}. The range of sulphate values used in the Table evidently covers most of the observed data. Although it might have been logical to vary SO_2 with the concentrations of sulphate, dissolved SO_2 itself only slightly affects pH (see Fig. 1), and we did not want to complicate the picture with further variations. Clearly a raining cloud low in both basic materials and L might be quite acidic. It is not possible to construct a frequency-of-occurrence histogram from these computations, but the above discussion indicates that pH values well below 5 might occur in pristine areas.

Considerable spatial and temporal variations in the acidity should be expected with lowest pH values occurring in situations with high concentrations of non-sea spray sulphate in aerosols, low concentrations of basic constituents (such as ammonia gas, CaCO_3) and low liquid water content in the clouds. As the sources of these basic constituents are mainly in continental areas whereas substantial emissions of sulphur compounds take place over the oceans, rain is likely to be more acidic in marine environments than over the continents. It is therefore not appropriate to use $\text{pH} = 5.6$ as a reference value against which human influences should be judged. We believe that, generally, pH alone is a poor indicator of departure from the natural

composition of rain and cloud water, mainly because of its large natural variability and the difficulty of measuring it accurately.

To assess man-made acidification it is necessary to study in detail the cycles—particularly of sulphur and nitrogen—that influence the acidity of atmospheric water, in different regions of the world, and the degree to which the cycles have been affected. Long-term (decades) programmes for monitoring the chemical composition of air and precipitation should be supplemented by direct estimates of emission fluxes from different types of sources potentially affecting the monitoring sites. A large number of variables must be studied simultaneously to define clearly natural and perturbed conditions.

Note that the present pH values in precipitation in the heavily industrialized regions are significantly affected by anthropogenic emissions. In Northern Europe and eastern parts of North America^{4,30,31}, annual average pH values of 4.3 are common. At a typical station in rural parts of southern Sweden (Forshult; lat. 60.15 N long. 13.78 E) 50% of the monthly precipitation samples during the period 1975 to 1980 had a pH value in the range 4.1–4.6, due primarily to H_2SO_4 and secondarily to HNO_3 (ref. 32). Annual sulphur deposition as SO_4^{2-} is typically 2–3 g m^{-2} , resulting in substantial ecological effects³¹. Clearly the sulphur and nitrogen cycles in these regions are much perturbed^{15,21} and if all industrial emissions of SO_2 and NO_x were stopped—emissions of NH_3 remaining the same—the pH would increase to > 5 . Note that low pH values reported for rainwater in remote locations^{33,34} need not be interpreted as due exclusively by natural processes. The possibility of long-range as well as local anthropogenic influences should be investigated. However, natural factors exist which can cause variable and significant decreases of pH from that expected from pure carbonic acid equilibrium.

Contribution No. 448 from International Meteorological Institute, Stockholm, Sweden.

Received 23 October; accepted 16 December 1981.

- Garrels, R. M. & MacKenzie, F. T. *Evolution of Sedimentary Rocks* (Norton, New York, 1971).
- Galloway, J. N., Likens, G. E. & Edgerton, E. S. *Science* **194**, 722–724 (1976).
- Likens, G. E. & Bormann, F. H. *Science* **184**, 1176–1179 (1974).
- Likens, G. E., Bormann, F. H. & Johnson, N. M. in *Some Perspectives of the Major Biogeochemical Cycles* SCOPE Rep. 17 (ed. Likens, G. E.) (Wiley, Chichester, 1981).
- Hallberg, R. O. in *Nitrogen, Phosphorus and Sulphur—Global Cycles* SCOPE Rep. 7, (eds Svensson, B. H. & Söderlund, R.) 93–101 (Swedish Natural Science Research Council, 1976).
- Granat, L. in *Nitrogen, Phosphorus and Sulphur—Global Cycles* SCOPE Rep. 7 (eds Svensson, B. H. & Söderlund, R.) 102–122 (1976).
- Nguyen, B. C., Gaudry, A., Bonsang, B. & Lambert, G. *Nature* **275**, 637–639 (1978).
- Hansen, M. H., Ingvorsen, K. & Jørgensen, B. B. *Limnol. Oceanogr.* **23**, 68–76 (1978).
- Cadle, R. D. *J. geophys. Res.* **80**, 1650–1652 (1975).
- Friend, J. P. in *Chemistry of the Lower Atmosphere* (ed. Rasool, S. I.) 177–201 (Plenum, New York, 1973).
- Cullis, C. F. & Hirschler, M. M. *Atmos. Envir.* **14**, 1263–1278 (1980).
- Rodhe, H. & Isaksen, I. *J. geophys. Res.* **85**, 7401–7409 (1980).
- Eriksson, E. *Tellus* **12**, 63–109 (1960).
- Rodhe, H. *Atmos. Envir.* **12**, 671–680 (1978).
- Rodhe, H. in *Nitrogen, Phosphorus and Sulphur—Global Cycles* SCOPE Rep. 7 (eds Svensson, B. H. & Söderlund, R.) 123–128 (Swedish Natural Science Research Council, 1976).
- Galloway, J. N. & Whelpdale, D. M. *Atmos. Envir.* **14**, 409–418 (1980).
- Taylor, G. S., Baker, M. B. & Charlson, R. J. in *Interaction of the Biogeochemical Cycles* (eds Bolin, B. & Cook, R.) (Wiley, London, 1982).
- Duce, R. A. *Biogeochemical Cycles and the Air/Sea Exchange of Aerosols* (Wiley, London, 1982).
- Sillén, L. G. in *Equilibrium Concepts in Natural Water Systems* Adv. Chem. Ser. no. 67, 45–56 (American Chemical Society, Washington DC, 1967).
- Junge, C. E. *Air Chemistry and Radioactivity* (Academic, New York, 1963).
- OECD. *The OECD Programme on Long Range Transport of Air Pollutants, Measurements and Findings* (Organization for Economic Cooperation and Development, Paris, 1977).
- Fletcher, N. H. *The Physics of Rain Clouds* (Cambridge University Press, 1962).
- Slinn, W. G. N. *et al. Atmos. Envir.* **12**, 2055–2087 (1978).
- Scott, B. C. *J. appl. Met.* **20**, 619–625 (1981).
- Maroulis, P. J., Torres, A. L., Goldberg, A. B. & Bandy, A. R. *J. geophys. Res.* **85**, 7345–7349 (1980).
- Garland, J. A. *Q. J. R. met. Soc.* **97**, 483 (1971).
- Lawson, D. R. & Winchester, J. W. *Science* **205**, 1267–1269 (1979).
- Huebert, B. J. & Lazrus, A. L. *J. geophys. Res.* **85**, 7337–7344 (1980).
- Bonsang, B., Nguyen, B. C., Gaudry, A. & Lambert, G. *J. geophys. Res.* **85**, 7410–7416 (1980).
- Bolin, B. *et al. Air Pollution Across National Boundaries; The Impact on the Environment of Sulfur in Air and Precipitation* (Royal Ministry for Foreign Affairs, Stockholm, 1971).
- Drablos, D. & Tollan, A. *Ecological Impact of Acid Precipitation*; Proceedings of an International Conference (SNSF, Ås, Norway, 1980).
- Granat, L. *Atmos. Envir.* **12**, 413–424 (1978).
- NOAA. *Geophysical Monitoring for Climatic Change* No. 7 (National Oceanic and Atmospheric Administration, Boulder, Colorado, 1979).
- Kerr, R. A. *Science* **212**, 1014 (1981).
- Larsen, T. V. & Harrison, H. *Atmos. Envir.* **11**, 1133–1141 (1977).

Mesozoic magnetic anomalies in the southern Natal Valley

S. W. Goodlad*, A. K. Martin* & C. J. H. Hartnady†

* Marine Geoscience Unit and † Precambrian Research Unit, Department of Geology, University of Cape Town, Rondebosch, South Africa

The recognition of Mesozoic magnetic anomalies M0–M12 within the submarine Natal Valley (south-west Indian Ocean) provides a further constraint on the relative motion between South America (Falkland Plateau) and Africa during the early opening phase of West Gondwanaland (Fig. 1). Colinear with M12, magnetic anomaly G (ref. 1) is delineated north of the South Tugela Ridge², an acoustic basement high which is suspected to have been emplaced within continental crust immediately north of the continental–oceanic crust boundary (COB) during the initial rifting phase. Starting from a revised West Gondwana reconstruction³ and using palaeo-poles of rotation^{1,3}, the Falkland Plateau can be rotated into the Natal Valley superposing the corresponding Georgia Basin and Natal Valley magnetic anomalies. This achieves an optimum continuity of inter-continental structural lineaments³ by overlapping of an anomaly intermediate to M10 and M11 (~M10N⁴). This suggests that the Natal Valley COB is located between anomalies M10 and M11, implying that in this zone of south-west Gondwana, seafloor spreading commenced at ~125 Myr.

Previous confirmation of the continental nature of the Falkland Plateau^{5–7}, necessitates its inclusion in recent reconstructions of south-west Gondwana^{3,7,8}. In these reconstructions it fits into the gap between south-east Africa and the Mozambique Ridge (Fig. 1). During the West Gondwana drift phase,

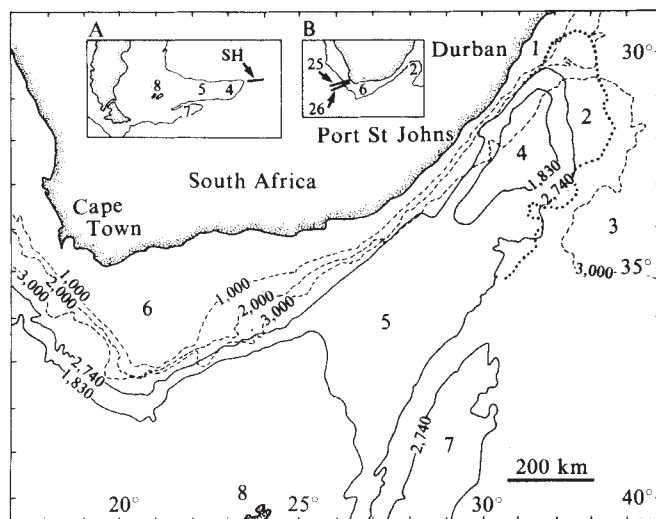


Fig. 1 Reconstruction of south-west Gondwana after Martin *et al.*³ depicting the close contiguity of the Falkland Plateau and Natal Valley physiographic features. Dashed line, South African bathymetry^{2,17} (1,000, 2,000 and 3,000 m). Solid line, South American/Falkland Plateau bathymetry¹⁸ (1,830 and 2,740 m); conversion factor 1 fathom = 1.83 m. Alternative position for north-east apex of the Falkland Plateau (2,740 m isobath) in reconstruction using Rabinowitz and LaBrecque¹ rotation parameters dotted. Physiographic zones: 1, Tugela Cone; 2, Natal Valley; 3, Mozambique Ridge; 4, Maurice Ewing Bank; 5, Falkland Plateau Basin; 6, Agulhas Bank (Outeniqua Basin); 7, Burdwood Bank; 8, Falkland Islands. A, The location of magnetic profile Shackleton¹¹ (SH) in the Georgia Basin east of the Falkland Plateau apex. B, Locations of magnetic profiles 25 and 26 (ref. 1) from the Cape Basin, south-west of South Africa.

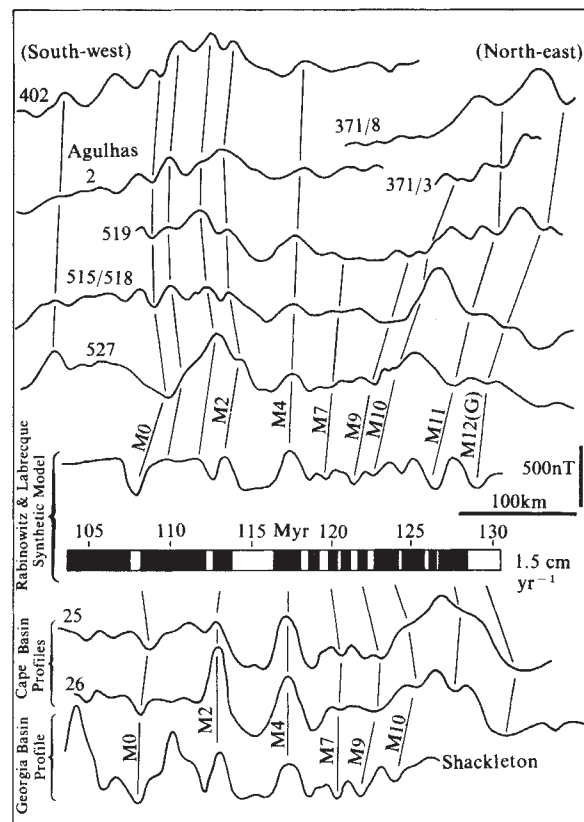


Fig. 2 Natal Valley magnetic profiles geographically aligned and correlated with respect to the magnetic block model of Rabinowitz and LaBrecque¹. The model was generated by using a half spreading rate of 1.5 cm yr⁻¹ for the Mesozoic reversal time scale of Larson and Hilde⁴. Profiles 402 (PR 2a), 371/8, 371/3, Agulhas 2, 519 (PR 3), 515/518 (PR 4) and 527 (PR 5) trend south-west–north-east within the Natal Valley sub-parallel to the Agulhas Fracture Zone. Notation in brackets refers to profile numbers used by du Plessis¹⁰. Profiles 25 and 26 (after ref. 1) from the Cape Basin and profile Shackleton (after ref. 11) from the Georgia Basin (immediately east of the Falkland Plateau) are included to show their general similarity to the Natal Valley traverses. Profile Shackleton is depicted as a mirror-image. Locations of profiles 25, 26 and Shackleton are shown in Fig. 1 insets.

the Falkland Plateau moved past Africa constrained by the orientation of the Agulhas/Falkland Fracture Zone^{9,10}. Recognition of magnetic anomalies on conjugate margins of the southern South Atlantic has helped to define the seafloor spreading history of South America and Africa¹ north of the Agulhas/Falkland Fracture Zone. In contrast, to the south of this fracture zone, the Natal Valley has been largely ignored as a prospective source of geophysical data for testing reconstruction of West Gondwana. The distant location (~40°) from the early pole of rotation^{1,3} suggests that seafloor spreading in the Natal Valley may have advanced at a fast rate and thus resultant magnetic anomalies should be widely spaced and easy to recognize. Magnetic anomaly lineations present in the Natal Valley should be aligned approximately perpendicular to the Agulhas Fracture Zone.

Four previously published magnetic traverses¹⁰ reassessed in conjunction with newly acquired data reveal a Mesozoic magnetic anomaly lineation pattern (Cape Sequence) within the southern Natal Valley (Fig. 2). Du Plessis¹⁰ recognized magnetic lineations normal to the coast but did not correlate these with any published Mesozoic anomaly models^{1,4}. The reinterpreted profiles (527, 515/518, 519, 371/3, 371/8, Agulhas 2 and 402) are geographically aligned and compared with the Mesozoic anomaly sequence of Rabinowitz and LaBrecque¹ in Fig. 2. Inter-profile correlation and correspondence to the synthetic model is good. The slight disparity of profile 527 with respect to

the synthetic is suspected to be due to navigational inaccuracy. The Natal Valley and Cape Basin magnetic signatures parallel to the spreading direction show similarities (Fig. 2), confirming correct anomaly sequence identification.

We recognize anomalies M12 to M0 with correlation being particularly good for M4 to M0 (Fig. 2). Anomaly M4 is prominent in all profiles as a broad positive anomaly 250 nT in amplitude while the M2–M0 composite anomaly is conspicuous as a positive triple peak some 200–500 nT in amplitude. Lineations defining M11, M9 and M7 are less distinct but still considered valid (Fig. 2). However, identification of the Mesozoic Cape Sequence rests primarily on recognition of anomalies M4–M0.

The observed profiles provide a reasonable fit to the model¹ presented in Fig. 2 and thus 1.5 cm yr^{-1} is proposed as an average Mesozoic half-spreading rate for the Natal Valley. In detail, however, the M4–M0 rate appears slightly slower than the 1.5 cm yr^{-1} model (the anomalies are more compressed) while a slightly faster spreading rate ($\sim 1.7 \text{ cm yr}^{-1}$) may be appropriate to the M11–M4 sequence. These velocities are comparable with those calculated for crust immediately east of the Falkland Plateau in the Georgia Basin¹¹.

An areal plot of the magnetic traverses (Fig. 3) show the anomalies to be in sub-parallel alignment with a strike trend normal to the Agulhas Fracture Zone (defined here by the Cape Slope Anomaly¹²). Oceanic crust that was created directly east of the Falkland Plateau in the Georgia Basin should theoretically generate mirror images of Natal Valley magnetic profiles although these may now be distorted by differing ambient field declination, inclination and skewness. A profile from the Georgia Basin¹¹, plotted in reverse in Fig. 2, exhibits a degree of similarity to Natal Valley anomalies.

Recently, Martin *et al.*³ proposed a revised fit of South America and South Central Africa (Fig. 1) primarily based on

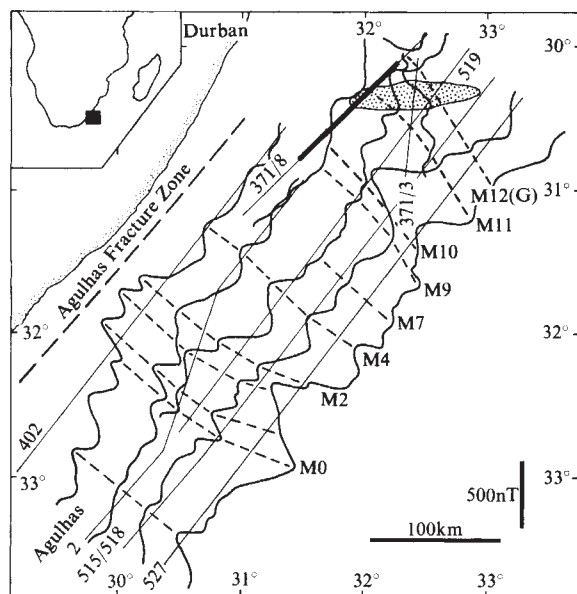


Fig. 3 Magnetic anomaly profiles plotted perpendicular to ship's tracks. Lines 402, 515/518, 519, 527, 371/3 and 371/8 were run between 1971 and 1977 on the RV *Thomas B. Davie*. Track Agulhas 2 was run on the MV *S A Agulhas* in August 1978. Inter-profile correlation identifying the Mesozoic anomaly sequence M12 (G) to M0 picked out by short dashed lines. Disparity of traverse 527 is suspected to be due to navigational inaccuracy. Note that the anomaly sequence trends normal to the Agulhas Fracture Zone (heavy dashed line adjacent to the coast). Enhanced portion of traverse 371/8 indicates location of the continuous seismic reflection profile depicted in Fig. 4. Dotted zone demarcates the seismically-defined extent of the South Tugela Ridge. Inset shows the study area in relation to southern Africa and Madagascar.

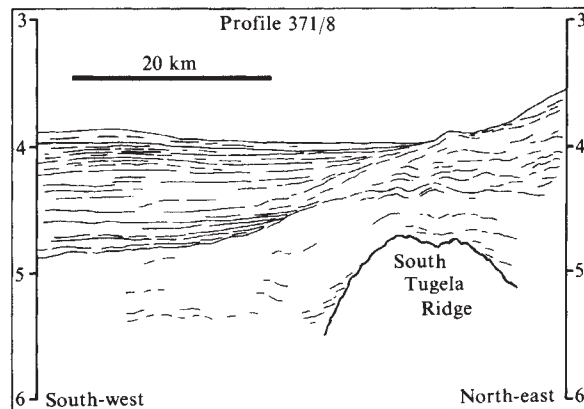


Fig. 4 Line tracing of continuous seismic reflection profile (371/8) showing the acoustic signature of the South Tugela Ridge. Profile location is shown in Fig. 3. Vertical scale is in seconds of two-way travel time.

cross-continent alignment of three major pre-drift tectonic features: (1) the eastern and western boundaries of the submarine Jurassic Outeniqua Basin (Agulhas Bank) and the Falkland Plateau Basin; (2) the northern tectonic limit of the Triassic Cape Fold Belt (South Africa) and a closely contiguous major morphological feature on the Falkland Plateau; and (3) the late Precambrian transcurrent fault and mylonite belts of Pernambuco (Brazil) and Fouban (West Africa).

This revised reconstruction³ (total pole of rotation 46.75° N , 32.65° W ; rotation angle 56.4°) is close to the refit of Rabinowitz and LaBrecque¹ but does not so tightly juxtapose conjugate margins of the Natal Valley and Falkland Plateau (see Fig. 1). This looser reconstruction³ therefore removes suspected continental basement overlap in the Natal Valley and Precambrian basement overlap in the Gulf of Benin (West Africa)³.

Seismic profiling surveys have delineated an approximately east–west oriented, deeply buried, basement ridge—the South Tugela Ridge (Fig. 4) underlying or immediately south of the steep Tugela Cone southern flank. The ridge location is depicted in Fig. 3. Significantly, in the revised reconstruction of Martin *et al.*³, the South Tugela Ridge becomes juxtaposed to the continental–oceanic crust boundary (COB) of the Falkland Plateau as approximated by the 2,740 m isobath on seismic refraction results⁵. Independent geological–geophysical evidence therefore suggests that the South Tugela Ridge may be situated just north of the Natal Valley COB and is probably a result of initial rift volcanics intercalating with splintered continental blocks.

Within the Natal Valley, the oldest correlative magnetic anomaly is identified as M12 (Figs 2 and 3) and corresponds in character to anomaly G of Rabinowitz and LaBrecque^{1,13}. Magnetic anomaly G, first mapped along the west coast of Africa^{1,13}, has been interpreted as a magnetic edge effect separating continental from oceanic basement. Reinforced by isostatic gravity modelling, Rabinowitz and LaBrecque use anomaly G to demarcate the COB. However, the west coast Africa magnetic lineation G, to which Rabinowitz and LaBrecque¹ assign an M12 age, has previously been dated as M13 by Larson and Ladd¹⁴ using the same raw data; thus dating of this anomaly is not unambiguous.

The revised fit of Martin *et al.*³ suggests that the south-west Africa COB may be sited $\sim 40 \text{ km}$ seaward of magnetic anomaly G, but still positioned within the landward gradient of the isostatic gravity anomaly. On the basis of reconstruction evidence³, the Natal Valley COB is suspected to be located south of the South Tugela Ridge implying that anomaly G is placed on continental crust.

Utilizing the new palaeo-spreading pole of Martin *et al.*³, the rotation of South America towards Africa achieves precise juxtaposition of Mesozoic anomalies (M0, M2, M4 and M10) in

areas both south and north of the Agulhas/Falkland Fracture Zone. Thus, using only one set of rotation parameters, abutment of Cape and Argentine Basin anomalies¹ is simultaneous to abutment of the equivalent Natal Valley and Georgia Basin anomalies¹¹. Complete kinematic description of the early spreading history will be more fully considered elsewhere. However, back-rotation to M10N (an anomaly intermediate to M10 and M11)⁴ accomplishes optimum contiguity of the inter-continental structural lineaments previously described. We therefore postulate that opening of the South Atlantic commenced at ~125 Myr (M10N using the time scale of Larson and Hilde⁴).

Accepting this looser reconstruction as compared to that of Rabinowitz and LaBrecque¹ (see Fig. 1), it is necessary to explain the existence of recognizable magnetic lineaments (M11 and M12) landward of M10N (approximate COB) as demarcated in Figs 2 and 3. A possible explanation may be that before final separation, crustal rifting allowed extensive intrusion of volcanic lavas which were magnetized in the Earth's ambient field. This mechanism, with possible interference from basement topography and non-magnetized slivers of continental crust, may explain the poorer M11 and M12 correlation with the modelled profile (Fig. 2) and severe anomaly distension on several profiles. Larson and Ladd¹⁴ likewise attribute poor correlation between anomalies older than M7 along the south-west Africa continental margin to disruption of the lineation pattern by the initial rifting phenomenon. Basement topography certainly induces modulation of the ambient field in this zone with high amplitude positive anomalies plotting over the South Tugela Ridge (Fig. 3, profiles 371/8, 371/3 and 519).

The collection of multichannel seismic data off the continental margin of south-west Africa^{15,16} has cast additional doubt on the 'anomaly G' refit¹ of South America and Africa. In particular, Uchupi and Austin¹⁵ locate their southwestern Africa COB at or closely seaward of the 3,000 m isobath. This placement is in general agreement with the reconstruction of Martin *et al.*³ which likewise corroborates the new magnetic evidence presented here. Although the south-west Africa COB of Gerrard and Smith¹⁶ is located even further seaward, constraints on data quality allow for some ambiguity in exact positioning (I. Gerrard, personal communication, 1981). These new multichannel data also imply that extensive intracontinental stretching took place along the passive margin of south-west Africa¹⁵ during the early opening phase of the South Atlantic. However, the reconstruction of Martin *et al.*³ in combination with the magnetic data outlined in this study suggests that post-breakup continental crust stretching may not have been active over the Falkland Plateau/Natal Valley zone.

We thank R. V. Dingle for a critical review, D. Murray and the CSIR for access to unpublished Agulhas 2 magnetic data, the Director of the Geological Survey for funding the project and permission to publish, and Mrs M. C. Hollander for typing the script.

Received 30 April; accepted 30 October 1981.

1. Rabinowitz, P. D. & LaBrecque, J. L. *J. geophys. Res.* **84**, 5973–6002 (1979).
2. Goodlad, S. W. thesis, Univ. Cape Town (in preparation).
3. Martin, A. K., Hartnady, C. J. H. & Goodlad, S. W. *Earth planet. Sci. Lett.* **54**, 293–305 (1981).
4. Larson, R. L. & Hilde, T. W. C. *J. geophys. Res.* **80**, 2586–2594 (1975).
5. Ewing, J. I., Ludwig, W. J., Ewing, M. & Eittreim, S. C. *J. geophys. Res.* **76**, 7118–7137 (1971).
6. Rabinowitz, P. D., Cande, S. C. & LaBrecque, J. L. *Ann. Acad. Brasil. Cienc.* **48**, 241–251 (1976).
7. Barker, P. F. *et al. Init. Rep. DSDP Leg 36*, 933–1014 (1977).
8. Norton, I. O. & Sclater, J. G. *J. geophys. Res.* **84**, 6803–6830 (1979).
9. Francheteau, J. & Le Pichon, X. *Bull. Am. Ass. petrol. Geol.* **56**, 991–1007 (1972).
10. du Plessis, A. *Nature* **270**, 719–721 (1977).
11. Barker, P. F. *Geophys. J.R. astr. Soc.* **59**, 131–145 (1979).
12. du Plessis, A. & Simpson, E. S. W. *Mar. geophys. Res.* **2**, 99–110 (1974).
13. Rabinowitz, P. D. *Bull. geol. Soc. Am.* **87**, 1643–1653 (1976).
14. Larson, R. L. & Ladd, J. W. *Nature* **246**, 209–212 (1973).
15. Uchupi, E. & Austin, J. A. *EOS* **62**, 407 (1981).
16. Gerrard, I. & Smith, G. C. *Paper presented at the Hedberg Research Conf. on Continental Margins*, Galveston (1981).
17. Simpson, E. S. W. *Geol. Soc. S. Afr.* **80**, Annexure 1–15 (1977).
18. Lonardi, A. G. & Ewing, M. *Phys. Chem. Earth* **8**, 79–121 (1971).

Magnetic granulometry results for intrusive rock samples

G. S. Murthy & R. R. Pätzold

Geomagnetic Research Laboratory, Physics Department, Memorial University of Newfoundland, St John's, Newfoundland, Canada A1B 3X7

'Magnetic granulometry' techniques have been developed^{1–4} to infer the domain state of magnetic minerals contained in rocks. Such techniques include study of the development of hysteresis of samples in fields of the order of 10 Oe (Rayleigh loops), study of hysteresis in high fields both at room temperature and at liquid nitrogen temperatures, and study of variation of magnetic susceptibility with temperature from –196 °C to 700 °C. These methods enable the magnetic properties to be quickly determined and so give an idea of the bulk magnetic behaviour of the rock sample. Although the usefulness of such methods for palaeomagnetic studies is limited because the remanent magnetic behaviour can be quite distinct from the bulk magnetic behaviour⁵, in general, the granulometry method is a good supplement to regular experimental methods in palaeomagnetism. The study of several thousands of basalt samples by this method^{6,7} has revealed insights into the magnetic behaviour of rocks. However, as part of the interpretive procedure of the method, the presence of low-field hysteresis loops has been considered a diagnostic feature of predominant superparamagnetic bulk behaviour of the rock sample. As we report here, we have observed, for the first time, low-field hysteresis loops associated with confirmed multi-domain magnetic grains, a possibility originally suggested on theoretical grounds by Néel⁸.

Although the interpretation of results from granulometric experiments is controversial⁷, there is agreement on the following basic points:

(1) A sample with multi-domain pure magnetite as the magnetic mineral will generally show $J_r/J_{max} \sim 0.2$ with approximate doubling of coercivity as the temperature of the sample is changed from room temperature (25 °C) to liquid nitrogen temperature (–196 °C), a phenomenon consistent with earlier work^{9,10}. As the magnetocrystalline anisotropy is more dominant than the shape anisotropy for multi-domain magnetic grains, magnetic susceptibility for this class of samples will be a function of the magnetocrystalline anisotropy constants K_1 and K_2 . Because K_1 and K_2 change sign at approximately –150 °C (refs 11, 12), an isotropic point with resultant susceptibility peak occurs close to that temperature. Low-field hysteresis loops for this class of samples have not previously been observed, and Fig. 1a, d, g shows their characteristics.

(2) A sample with single-domain magnetite as the magnetic mineral will have $J_r/J_{max} = 0.5$ at room temperature, increasing to 0.8 at –196 °C, with a corresponding increase in coercivity. For this class of samples, due to the dominance of shape anisotropy over magnetocrystalline and/or stress anisotropy, the susceptibility peak at the isotropic point would be suppressed and a monotonic susceptibility decrease is usually observed as the temperature is lowered from 25 °C to –196 °C. This class of samples does not exhibit Rayleigh loops. Typical results are shown in Fig. 1b, e, h.

(3) In the class III samples, which were interpreted as being due to the predominantly superparamagnetic bulk behaviour of the sample, Rayleigh loops would be observed^{13,14}, due to the interaction of clusters of superparamagnetic grains. The high-field hysteresis would show $J_r/J_{max} \sim 0.2$ at 25 °C, increasing to >0.80 at –196 °C, with a substantial increase in coercivity due to grains which are superparamagnetic at room temperature becoming effectively single-domain at –196 °C because of the lowering of the thermal energy. The susceptibility-temperature

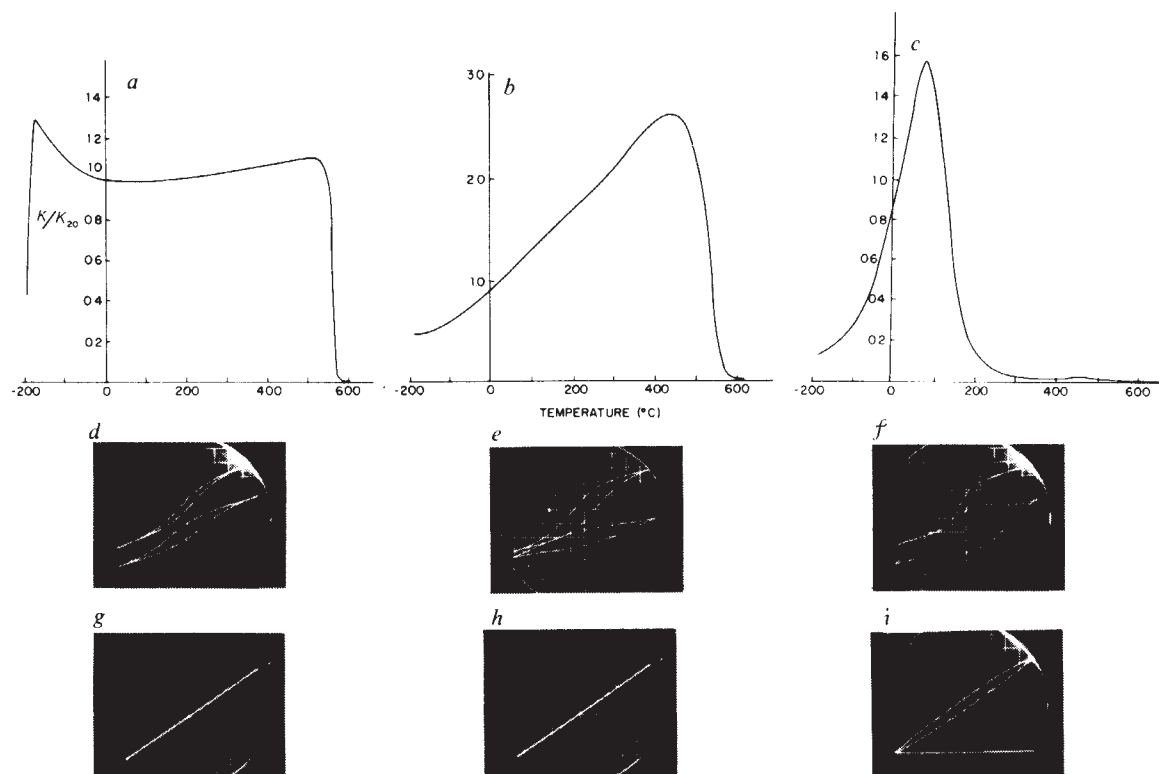


Fig. 1 Basis for interpretation of magnetic granulometry. *a-c*, Normalized susceptibility versus temperature; *d-f*, magnetic hysteresis in a field of 1,200 Oe at room temperature (upper trace) and at -196°C (bottom trace); *g-i*, Rayleigh loops in a field of 10 Oe. *a, d, g* are typical responses for class I behaviour (multi-domain) as obtained on G215C, a dolerite dyke sample from Newfoundland; *b, e, h* are typical responses for class II behaviour (single domain) as obtained on a basalt sample from a DSDP drill hole; *c, f, i* are typical responses for class III behaviour (superparamagnetic) as obtained on a basalt sample from a DSDP drill hole.

curve would show a decrease in normalized susceptibility to ~ 0.2 at -196°C and a location of the peak which differs in different samples, but which is nearly always at temperatures $< 300^{\circ}\text{C}$, followed by a long tail. Figure 1*c, f, i* shows typical results exhibited by a member of this class of samples. Class III and class II behaviours were previously designated type 1 and type 2^{15,16}.

The most controversial aspect of this interpretation is the inference of superparamagnetic grains from the presence of Rayleigh loops and the typical susceptibility-temperature curves of the class III type. The observed grain sizes in most rocks exhibiting this behaviour are of the order of hundreds of micrometres, and for the grains to behave like superparamagnetics, they must be effectively subdivided into SP sizes. The question arises of why the observed Rayleigh loops are not due to multi-domain grains. Néel⁸ alluded to the possibility of multi-domain grains exhibiting hysteresis in low fields. The BH_c/A parameter (where B and A are the quadratic and linear coefficients of Rayleigh relation and H_c is the bulk coercive force) for such loops will be of the order of 0.05, while the value for this quantity for the observed loops^{13,15} is of the order of 10. The reason cited by Néel for the smaller BH_c/A for multi-domains is that the demagnetizing factor is to be taken into account for multi-domains. In all the thousands of measurements conducted to date, all confined to basalt samples, a Rayleigh loop was never found associated with $K-T$ curves and hysteresis loops of the class I type. Due to the observed association of the susceptibility-temperature variation for the multi-domain samples, well established in theory and experiment, with the corresponding 'absence' of Rayleigh loops for such samples, the established granulometric interpretation was that Rayleigh loops exhibited by multi-domain samples cannot be seen. In particular, it was noted¹⁵ that on the present Rayleigh loop tracer, a multi-domain Rayleigh loop, with its low BH_c/A value, would not be wider than the thickness of the oscilloscope trace itself.

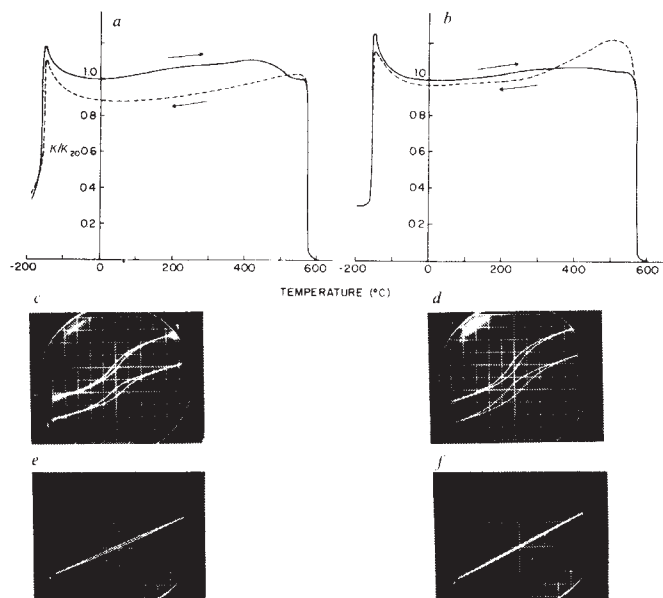


Fig. 2 Magnetic granulometry results: *a, c, e*, for GR353C, a dolerite dyke sample from the Grenville Province of the Canadian Shield; *b, d, f*, for N-69.1, an anorthosite sample from Nain, Labrador; *a, b*, normalized susceptibility versus temperature; *c, d*, magnetic hysteresis in a field of 1,200 Oe at 25°C (upper trace) and at -196°C (bottom trace); *e, f*, Rayleigh loops in a field of 10 Oe.

We extended the granulometric studies to older intrusive rocks (dolerite dykes, anorthosites and gabbros). Samples studied include the Grenville dykes, Indian Head and Nain anorthosites, and gabbros from Newfoundland. Earlier palaeomagnetic studies of these rock units revealed that their remanent magnetism is extremely stable (Koengsberger ratios ~ 10) and

that their remanence, particularly for the anorthosites, resided in elongated magnetite grains exsolved within the silicates. The present systematic studies include: (1) magnetic susceptibility variation from -196°C to 700°C ; (2) hysteresis loops at -196°C and at room temperature in a field of 1,200 Oe (note that a 1,200-Oe field may not always result in saturation of the sample); (3) Rayleigh loops in a field of 10 Oe.

Most samples are class I type; class II and III types are virtually absent. This confirms the earlier conclusions⁶ that most 'old' rocks are predominantly multi-domain in nature and that superparamagnetic behaviour is virtually absent in the 'old' rocks. However, many samples exhibited a new type of behaviour—a Rayleigh loop at low fields, associated with class I type hysteresis and K - T curves. Typical examples of results for this behaviour are shown in Fig. 2 for two samples, a dolerite dyke and an anorthosite. These samples do not fall into the three classes discussed earlier. Whereas the presence of Rayleigh loops suggests SP behaviour, the K - T curves and the high-field hysteresis strongly favour the multi-domain state. An explanation in terms of mixtures³ does not hold here, as experiments lead to different inferences according to the previous interpretation (Fig. 1). Further, an SP component in a mixture should have modified the K - T curve from what is now a 'pure' multi-domain curve. Therefore, there is no doubt that the bulk magnetic properties of these samples, including their Rayleigh loops (Fig. 2c, f) are caused by magnetite in the multi-domain state. The value of the BH_c/A parameter for these loops is similar to some of the lower values in previously observed loops inferred as being due to effectively superparamagnetic grains.

It is possible that the observed Rayleigh loops reported here are due to a highly conducting nature of the samples (C. Radhakrishnamurthy, personal communication) and the resulting eddy currents. Although the conductivity of these samples is of the same order as that of other intrusive rocks of the type studied by us, we tested for this possibility by crushing some of the samples to fine ($<50\text{ }\mu\text{m}$) sizes and remeasuring the Rayleigh loops. The Rayleigh loops were still present, although in slightly diminished thicknesses. We conclude that the observed Rayleigh loops for our samples are real and are not due to eddy current effects. This leads to the following possibilities. (1) With the K - T curve being the main diagnostic feature of multi-domain nature of a magnetic grain (class I type), two types of multi-domain behaviour exist: those samples that do not exhibit a Rayleigh loop and those that do show a Rayleigh loop. (2) Rayleigh loops associated with class III type K - T curves may be interpreted as due to magnetic grains being in a superparamagnetic state. (3) Exhibition of Rayleigh loops is a necessary condition for SP behaviour but is not a sufficient diagnostic feature. Further work is required to clarify these possibilities.

We thank Dr C. Radhakrishnamurthy for discussions and advice on various aspects of this observation, and Professor E. R. Deutsch for a critical review of the manuscript. This work was supported by a Natural Sciences and Engineering Research Council of Canada operating grant to G.S.M.

Received 14 July; accepted 30 December 1981.

1. Radhakrishnamurthy, C., Raja, P. K. S., Likhite, S. D. & Sahasrabudhe, P. W. *Pure appl. Geophys.* **93**, 129–140 (1972).
2. Radhakrishnamurthy, C. & Deutsch, E. R. *J. Geophys.* **40**, 453–465 (1974).
3. Radhakrishnamurthy, C., Likhite, S. D. & Sahasrabudhe, P. W. *Earth planet. Inter.* **13**, 289–300 (1977).
4. Radhakrishnamurthy, C., Deutsch, E. R. & Murthy, G. S. *J. Geophys.* **45**, 433–446 (1979).
5. Murthy, G. S., Pätzold, R. & Brown, C. *Phys. Earth planet. Inter.* **26**, 72–80 (1981).
6. Radhakrishnamurthy, C., Likhite, S. D., Deutsch, E. R. & Murthy, G. S. *Proc. Ind. Acad. Sci.* **A87**, 235–243 (1978).
7. Senanayake, W. E. & McElhinny, M. W. *Phys. Earth planet. Inter.* **26**, 47–55 (1981).
8. Neel, L. *Adv. Phys.* **4**, 191–243 (1955).
9. Morrish, A. H. & Watt, L. A. K. *J. appl. Phys.* **29**, 1029–1033 (1958).
10. Bean, C. P. *J. appl. Phys.* **26**, 1381–1383 (1955).
11. Bickford, L. R. *Phys. Rev.* **78**, 449–457 (1950).
12. Syono, Y. & Ishikawa, Y. *J. phys. Soc. Jap.* **18**, 1230–1232 (1963).
13. Radhakrishnamurthy, C. & Sastry, N. P. *Proc. Ind. Acad. Sci.* **A72**, 94–102 (1970).
14. Néel, L. *C. r. hebdo. Séanc. Acad. Sci., Paris* **270**, 1125–1130 (1970).
15. Deutsch, E. R. & Pätzold, R. *J. geophys. Res.* **81**, 4188–4198 (1976).
16. Dunlop, D. J. & Hale, C. J. *Can. J. Earth Sci.* **14**, 716–744 (1977).

Explicit dialogue between left and right half-systems of split brains

D. M. MacKay & Valerie MacKay

Department of Communication and Neuroscience,
University of Keele, Keele, Staffordshire ST5 5BG, UK

Much debate has surrounded the interpretation of the well known 'split brain' syndrome, produced by section of the corpus callosum in man [for useful surveys of early studies of the effects of callosal damage (for example by Liepmann (1900), Goldstein (1908) and Bonhoeffer (1914)) see refs 1–3]. On the basis of the observation that left and right half-systems can find themselves in conflict when disparate instructions are flashed to left and right visual hemifields, Sperry⁴ has described such patients as having "two free wills inside the same cranial vault". MacKay^{5–9}, among others, has argued for a more parsimonious interpretation of the available evidence, suggesting that it demonstrates duality only at the 'executive' level where continuous goal-pursuit is administered, rather than at the 'normative' supervisory level where the agent's priorities and criteria of evaluation are themselves subject to evaluation and revision. We report here some experiments in which a split-brain patient showed considerable capacity for explicit interhemispheric exchange of information and instructions, yet gave no sign of the duality of supervisory initiative that might have been expected on the hypothesis of 'two free wills'.

Gazzaniga and co-workers^{6,10} have shown both in monkey and man that each hemisphere can apply disparate criteria of evaluation to the same stimulus, as well as expressing disparate preferences in the same situation. This indicates convincingly that both evaluative and organizing systems are split at the executive level, but it falls short of proof that the brain in such patients embodies two independent systems for evaluating priorities themselves—the characteristic human activity with which the term 'will' is associated.

One relevant question is whether, and if so to what extent, one 'half-system' in these cases can be induced to demonstrate any recognition of the other half-system as 'another person'. Through the generous cooperation of R. W. Sperry and M. S. Gazzaniga, we have recently contrived a number of laboratory situations designed to give systematic opportunities for such 'interpersonal' types of interaction. Some of these have been described elsewhere⁹. In the case described here our strategy was to train each 'half' in turn to engage in interpersonal exchanges with an experimenter in a simple competitive guessing game, and then to encourage each to use its acquired symbolic skills to engage the other 'half' in the same manner. The question was whether we would see any signs of simultaneous conscious evaluation of priorities from disparate standpoints, such as would be shown by two conscious human agents in a bargaining situation.

The patient (J.W.) was a right-handed 27 yr-old man who had had his corpus callosum, but not his anterior commissure, sectioned in 1977 to help prevent the spread of epileptic seizures from one hemisphere to the other¹¹. Apart from side effects of anticonvulsant medication he was in good health, and cooperated with enthusiasm in the experiment. Preliminary tests confirmed observations by Sidtis *et al.*¹² that he was unable to report vocally any visual stimuli flashed to the left of his fixation point, although his left hand could correctly point to or retrieve objects represented by these stimuli. A single numeral randomly selected from the range 0 to 9 was flashed for 150 ms at an eccentricity of 4° to the left of the fixation point, simultaneously with a distractor (one or more letters of the alphabet) flashed at a similar eccentricity to the right. J.W. reported vocally that he saw only the distractor. Both vocal and right-manual responses were normal for stimuli flashed to the right half-field.

Each half-system in turn was then trained to participate (on opposite sides) in a guessing game of the '20 questions' type in dialogue with one experimenter (E).

(1) E first assumed the role of 'guesser', attempting to name a 'mystery number' written by the second experimenter on a piece of paper in free view of J.W. E invited J.W. to use his left hand to point to appropriate answers on a card also in free view, bearing the words 'go up', 'go down', 'OK'. The left hand proved immediately capable of providing reliable feedback in this mode.

(2) E then invited J.W. to "guess a number between 0 and 9 which I have been shown". In response to each guess by J.W., E pointed mutely to one of the three messages on the card. J.W. quickly learned to adapt his vocal guesses so as to home in on the correct answer under this mute feedback from E.

(3) Finally, single digits between 0 and 9 were flashed as before to J.W.'s left half-field, with distractors to the right; and E proposed that J.W. should fulfil both roles, using his mouth to make guesses and his left hand to offer feedback by way of the card. J.W. found no apparent difficulty in fulfilling this dual role, although it was notable that all his guesses were addressed to E, who invariably had to ask the left hand to answer. He proved well able to 'talk himself down' on to the correct target each time, frequently showing satisfaction by a smile when the target was finally named. A useful check that he was not just play-acting (which seemed in any case improbable) was provided by an episode when his mouth guessed '1' and his left hand pointed to 'go down'. J.W. complained vocally that there were "no more numbers down there", and had to be reminded of the possibility of '0' (which was the correct answer).

These results make clear that symbolic information-exchange by question and answer can be sustained between left and right half-systems of a split-brain patient. Various forms of cross-cueing have, of course, been known and studied in such patients for many years¹³. Sidtis *et al.*¹² have also shown that a '20 questions' guessing procedure on the part of an experimenter could help a patient with a partially sectioned corpus callosum to 'home in on' the name of a stimulus flashed to the left half-field; but it was by no means obvious that J.W.'s two half-systems could use a public symbolic system of this kind to engage in systematic dialogue with one another.

However, we are still left with the vexed question of whether the two half-systems here embodied 'two free wills' in the sense in which, say, two normal human players on opposite sides of a guessing game have independence of will. An ideal demonstration of such 'duality of will' would be the capacity to bargain: that is, to struggle to induce changes in one another's evaluative criteria or priorities as distinct from merely struggling physically in pursuit of conflicting goals^{8,9}.

Accordingly, E next proposed to J.W. that "the left hand (LH) should be paid by the right hand (RH) for each item of information". RH was provided with an open box containing a limited supply of tokens, and instructed to pay one token into a similar box on the left for each correct answer by LH. RH was rewarded (with additional tokens) in inverse proportion to the number of guesses required to identify each flashed numeral. Any occasional mistakes made by LH were penalized by repaying three tokens per mistake to RH.

As elaboration of the game in this way proved congenial to J.W., we next introduced the possibility of evaluative conflict by proposing to "ask the left hand whether it would like to be paid more for each answer". LH was asked to reply by pointing to a card representing one, two and three tokens; and it immediately indicated 'three tokens'. As this soon had the (foreseeable) effect of rendering RH bankrupt and halting the game, E then said "Let's ask the left hand whether it would settle for less". LH immediately pointed to 'two tokens'; but at the same time, and apparently as an integral part of the action, J.W.'s mouth said "Sure, make it two tokens". This was typical of the impression conveyed throughout this second phase of the experiment—that although J.W.'s LH and RH were substantially separate at the cognitive level, their priorities gave no evidence of being under the supervision of two independent normative systems⁷⁻⁹.

It is, of course, impossible to draw firm conclusions from such negative evidence. One can always explain absence of conflict by postulating suppressive (internal) influences from one half-system on normative processes in the other that would otherwise be independent; alternatively, it could be argued that the two half-systems perceived the rewards as jointly shared, and thus had no incentive to quarrel (M. S. Gazzaniga, personal communication). Our evidence is consistent, however, with the null hypothesis that at the higher levels of supervisory evaluation J.W. has not two half-systems but one undivided system, presumably associated with deep neural structures not surgically separated. This null hypothesis is strengthened by the well known fact^{5,6} that emotive stimuli to the mute hemisphere can elicit signs of corresponding emotion in the speaking subject. At least, the onus of proof seems now to rest on anyone who would claim that patients like J.W. have, nevertheless, 'two free wills'. Perhaps his own spontaneous comment, offered more than once, is illuminating: "Are you guys trying to make two people out of me?"

These experiments were possible only through the generous cooperation of Dr M. S. Gazzaniga and his colleagues, particularly Drs Jeffrey Holtzman, John Sidtis and Bruce Volpe, in the Division of Cognitive Neuroscience at Cornell University Medical College in New York City. We also thank J.W. for his willing participation. Financial support was received from a Sloan Foundation Grant to Dr Gazzaniga, from the University of Keele to D.M.M. and from the Wellcome Trust to V.M.

Received 20 October 1981; accepted 19 January 1982.

1. Geschwind, N. *Brain* **88**, 584-644 (1965).
2. Zangwill, O. L. in *Hemisphere Function in the Human Brain* (eds Dimond S. J. & Beaumont, J. G.) (Elek Science, London, 1974).
3. Joynt, R. J. in *Hemispheric Disconnection & Cerebral Function* (eds Kinsbourne, M. & Smith W. L.) (Thomas, Springfield, 1974).
4. Sperry, R. W. in *Brain & Conscious Experience* (ed. Eccles, J. C.) 293-313 (Springer, New York, 1966).
5. Gazzaniga, M. S. *The Bisected Brain* (Appleton-Century-Crofts, New York, 1970).
6. Gazzaniga, M. S. & LeDoux, J. E. *The Integrated Mind* (Plenum, New York, 1978).
7. MacKay, D. M. in *Brain & Conscious Experience* (ed. Eccles, J. C.) 312-313 (Springer, New York, 1966).
8. MacKay, D. M. in *Consciousness and the Physical World* (eds Josephson, B. D. & Ramachandran, V. S.) 95-113 (Pergamon, Oxford, 1980).
9. MacKay, D. M. *Neural Communication and Control* (eds Székely, G., Lábos, E. & Damjanovich, S.) 315-332 (Pergamon, Oxford and Akadémiai Kiadó, Budapest, 1981).
10. Gazzaniga, M. S. *Am. Scient.* **60**, 311-317 (1972).
11. Wilson, D. H., Reeves, A. & Gazzaniga, M. S. *Neurology* **28**, 649 (1978).
12. Sidtis, J. J., Volpe, B. T., Holtzman, J. D., Wilson, D. H. & Gazzaniga, M. S. *Science* **212**, 344-346 (1981).
13. Gazzaniga, M. S. *Scient. Am.* **217**, 24-29 (1967).

Somatotopic maps are disorganized in adult rodents treated neonatally with capsaicin

P. D. Wall*, M. Fitzgerald*, J. C. Nussbaumer†, H. Van der Loos† & M. Devor‡

* Cerebral Functions Research Group, Department of Anatomy, University College London, London WC1E 6BT, UK

† Institute of Anatomy, University of Lausanne, CH-1011 Lausanne, Switzerland

‡ Institute of Life Sciences, Hebrew University, Jerusalem, Israel

The somatosensory system is characterized at each stage of the projection pathways by the presence of maps formed by an orderly arrangement of cells and of the incoming fibres. Each cell responds to a definite area of skin, its receptive field (RF). If cell locations and their RFs are plotted side by side, they make up a continuous map of the body surface. Given the stability and repeatability of these maps, it was of interest to find that lesions which destroyed the input to one part of the map of the spinal cord in adults were followed by a readjustment of the RFs of cells which had lost their normal input¹. It was more surprising to discover in adult cats and rats that peripheral nerve lesions which do not produce an anatomical destruction of spinal cord afferents were followed by the appearance of novel RFs in cells

deprived of their normal physiological input^{2,3}. Capsaicin given to neonatal mice or rats destroys unmyelinated afferents. We report here that this is followed by an expansion of the normal receptive fields supplied by myelinated afferents^{2,3}. Our results suggest that unmyelinated fibres may be involved in controlling the connectivity of myelinated afferents with first central cells.

Experiments in rats show that the foot and toe representation is present in the medial two-thirds of the dorsal horn of segments L4 and L5 of the spinal cord. If this region is searched to a depth of 1 mm with fluid-filled microelectrodes in intact adult rats anaesthetized with Nembutal, an average of five cells responding to the foot or toes are isolated in each track. As the RFs remain stable over many hours of recording and since the maps are accurately repeated from animal to animal, it is reasonable to expect that abolition of a particular input would silence the cells which are normally dominated by that input. We therefore searched the same region with a total of 51 tracks in 13 animals in which the sciatic and saphenous nerves that supply the foot had been cut on the day of the experiment, and found only 0.1 responding cells per track; these few RFs were on the leg. The same virtual absence of responding cells was observed in six animals 2 and 3 days³ after the nerves had been cut. However, in 13 animals with a total of 69 tracks surviving 4–8 days after nerve section, the value rose to 3.2 cells per track. All these RFs were on the upper leg and pelvis and were supplied by intact nerves³. Similar results have been seen in the cat following peripheral nerve section². These new RFs in the foot region could not have been supplied by sprouts growing from the nerve sectioned 4–8 days earlier because resectioning of the nerve on the day of the experiment produced no change in the recordings.

Lesions of nerves in the hind legs, unlike root lesions, produce no frank degeneration of central afferent terminals in the spinal cord for at least 2 weeks after the lesion^{2,3}, but they do interrupt impulse and chemical transport from the periphery to the spinal cord. We suspected that the component of the peripheral nerve responsible for signalling to the spinal cord that the peripheral nerve was cut might be the unmyelinated C fibres, as anatomical and chemical changes have been reported in their spinal cord terminals and not in the terminals of myelinated afferents following peripheral nerve section^{4–6}.

Capsaicin, the hot compound in paprika, produces rapid, substantial and permanent destruction of C fibre sensory afferents in dorsal roots and in the trigeminal nerve if injected into neonates^{7–10}. This effect seems to be remarkably specific. Ultrastructural studies of cord and trigeminal nuclei show only degeneration of the terminals of fine afferents on laminae I and II^{11,12}. Immunocytochemistry reveals depletion of substance P only in dorsal horn, trigeminal nuclei, nerves and roots, with no depletion in ventral horn or elsewhere in the brain^{12–14}. The drug has no effect on enzymes and peptides intrinsic to the dorsal horn: neurotensin, glutamic acid decarboxylase and catecholamine amine transferase^{12,15}. Unlike capsaicin given systemically to adults, there is no evidence that neonatal capsaicin produces hypothalamic damage¹⁶. In our hands the peripheral lesion seems to be limited to the unmyelinated afferents⁹, but others have reported a small and variable loss of some myelinated fibres with high doses^{8,10}.

Litters of Sabra and Wistar rats under ether anaesthesia were injected with capsaicin (50 mg per kg intraperitoneally) dissolved in 10% Tween 80, 10% ethanol in saline on the second and fourth day of life. Littermates were injected with the vehicle. The animals were returned to their mothers and allowed to grow to adults weighing over 200 g. They were apparently healthy, with normal sleeping, eating and grooming behaviour and no deficit in motor activity.

The effectiveness of the drug was tested in three ways. First, the peripheral C compound action potential was measured by stimulating the dorsal root (5 mA, 500 μ s) and recording from the sural nerve at the beginning of each experiment. A C wave was never recorded in capsaicin-treated animals from dorsal root stimulation ($n = 4$) whereas in vehicle-treated animals and untreated controls a clear C compound action potential was

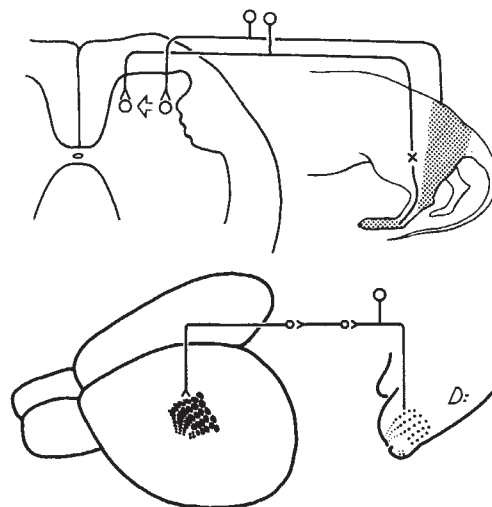


Fig. 1 Above: In the normal rat, the nerves from the foot and toes project to the medial part of the lumbar dorsal horn. The proximal parts of the leg project to the lateral cells in the dorsal horn. If the nerves to the foot are cut in normal animals, most of the medial cells fail to respond to peripheral stimulation for a period of 0–3 days after the cut. During the period 4–8 days after nerve section, many medial cells begin to respond to stimulation of proximal leg. In animals treated neonatally with capsaicin, medial cells respond to the upper leg immediately after peripheral nerves to the foot have been cut. Below: In the normal mouse, each individual large whisker projects by way of the trigeminal nerve, the trigeminal sensory nuclei and the thalamus to the cortex. In the parietal cortex there are barrels of cells in which each cell responds to the movement of one whisker only. Almost all responding cells in an individual barrel respond to the same single whisker. In mice treated with capsaicin soon after birth, the barrels and whiskers are present in their normal arrangement but no cell was encountered which responded to only one whisker. All recorded cells responded to 2–7 whiskers (average 3.8).

evoked. Second, the spinal cords were removed from four animals and stained for immunoreactive substance P. Substance P is normally highly concentrated in the region of small fibre terminals in the dorsal horn¹⁷ and in all the tested cords it was substantially depleted. Finally, cords from three rats were stained for fluoride-resistant acid phosphatase⁴. Again, there was substantial depletion of the enzyme from the dorsal horn terminals. Four vehicle control littermates were also tested by these three methods and the measurements were normal. It therefore appeared that the rats treated neonatally with capsaicin had lost a substantial number of their peripheral C fibres.

The spinal cord maps of 10 neonatally capsaicin-treated rats, 6 vehicle-treated and 4 untreated control animals were then examined. Low-thoracic spinalized preparations were used to increase the excitability of the cord and therefore reveal any aberrant cells in control animals. The rats were either anaesthetized with Nembutal or decerebrated under short-acting Althesin anaesthesia. The saphenous and sciatic nerves were cut acutely at knee level on the day of the experiment and the medial dorsal horn of segments L4 and L5 searched with electrode tracks for cells that responded to skin stimulation on the upper leg above the knee.

In normal and vehicle controls under Nembutal anaesthesia, 0.9 cells were found per track ($n = 43$). Decerebrate controls showed 1.8 responding cells per track ($n = 31$). This region was therefore clearly largely silenced by acute denervation of the lower limb, the actual number of responding cells being somewhat dependent on the excitability of the cord in different preparations. In capsaicin-treated animals with acute deafferentation of the foot, 2.6 cells per track ($n = 64$) were found in six Nembutal-low spinal animals. In the decerebrate preparations, the four treated animals had 6.7 cells per track ($n = 15$). The number of aberrant RFs in the foot area was significantly greater in the capsaicin-treated animals than in

controls ($P < 0.001$, t -test). The receptive fields were on the upper leg and in some cases (1.2 cells per track) extended to the tail and over to the contralateral limb. The responses were induced by firm pressure or pinch and often also by light touch. Thus, the capsaicin-treated animals have the same type of distorted somatotopic cord map after acute peripheral nerve section as occurs in normal rats after chronic nerve section.

The somatotopic map of spinal cord is not as detailed as that in the somatosensory cortex. A particularly precise and extensively investigated projection is that of the whisker follicles mapped on the parietal cortex of the mouse, in which an anatomical correlation is established between individual whiskers and discrete cytoarchitectonic units in the parietal neocortex, the barrels¹⁸. Furthermore, Van der Loos and Woolsey¹⁹ showed that limited neonatal lesions of particular whiskers resulted in failure of development of the corresponding barrels. It was later shown that the occasional appearance of an extra whisker is associated with an extra barrel in the topologically equivalent cortical locus²⁰. Following previous studies on cortical physiology of whisker responses^{21,22}, Nussbaumer (manuscript in preparation) has examined the details of the physiological aspect of this tight anatomical mapping in mice and has found an extraordinary specificity of both anatomical and physiological correlation between the periphery and its cortical representation. In penthane/nitrous oxide-anaesthetized mice, 117/121 tracks through layer IV of these cortical barrels revealed cells responding to only one whisker. In the remaining four tracks, the cells responded to two neighbouring whiskers. In view of this finding, it was obvious that we should test the effect of neonatal capsaicin. Two litters of mice (CBA/CA and ICR) were treated on the second and fifth day of life with 50 mg capsaicin per kg in the same manner as the rats. Littermates were treated with the vehicle alone. The animals were returned to their mothers, allowed to grow to adulthood and examined after more than 4 months.

In two vehicle-treated controls the cortex showed the same results as in normal mice in that five marked penetrations into layer IV of the barrelfield revealed cells responding to only one whisker. In the capsaicin-treated mice of both strains, the cortical anatomy as observed in Nissl preparations showed normal barrelfields. The size and distribution of whiskers were also normal. However, on penetrating the cortex, the recordings were strikingly abnormal. In two CBA and three ICR mice examined with a total of 18 penetrations, each penetration yielded one or more responding cells. No recording sites showed cells responding to only one whisker. The number of whiskers that could excite a cell varied from 2 to 7 (average 3.8). Invariably the whisker producing the most intense excitation was that which would normally have monopolized the barrel within which the recording was being made.

These experiments show that the RFs of some cells in rat spinal cord and in mouse cortex are greatly expanded in adults treated neonatally with capsaicin (Fig. 1). In spinal cord, it has been shown by electrical stimulation of the skin that many of the cells with expanded RFs are excited monosynaptically by myelinated afferents^{2,3}. The abnormal RFs of these cord cells may therefore be produced by abnormal coupling of myelinated afferents to central cells. For the mouse cortex cells with expanded RFs, a similar explanation would apply if trigeminal myelinated afferents establish abnormal contacts with cells in the trigeminal sensory nuclei.

These results suggest a testable hypothesis that the normal receptive fields of cord and medulla cells excited by myelinated sensory afferents require the presence of unmyelinated afferents. We propose that the effect of neonatal capsaicin is limited to its destruction of afferent fibres in the periphery. This is supported by the absence of signs of direct central destruction^{11,12,16}, and by results to be reported elsewhere²³ that local application of capsaicin to a single nerve in the adult also results in abnormal expanded receptive fields limited to cells supplied by the treated nerve. Although it is possible that the toxin has produced some subtle undetected change in the myelinated

afferents, in the case of single nerve lesions it is evident that the novel receptive fields are supplied by myelinated afferents which were not touched by the lesion. We therefore propose that unmyelinated afferents have a role in the establishment of ordered connections between myelinated afferents and the first central cells.

Received 17 November; accepted 22 December 1981.

1. Wall, P. D. *Phil. Trans. R. Soc. B* **278**, 361–372 (1977).
2. Devor, M. & Wall, P. D. *J. comp. Neurol.* **199**, 277–291 (1981).
3. Devor, M. & Wall, P. D. *J. Neurosci.* **1**, 679–684 (1981).
4. Devor, M. & Claman, D. *Brain Res.* **190**, 17–28 (1980).
5. Barbut, D., Polak, J. M. & Wall, P. D. *Brain Res.* **205**, 289–298 (1981).
6. Csillik, B. & Knyihar, E. *Prog. Neurobiol.* **10**, 203–230 (1978).
7. Jancso, G., Kiraly, E. & Jancso-Gabor, A. *Nature* **270**, 741–743 (1977).
8. Lawson, S. N. & Nickels, S. M. *J. Physiol., Lond.* **303**, 12P (1980).
9. Scadding, J. W. *J. Anat.* **131**, 473–484 (1980).
10. Nagy, J. I., Hunt, S. P., Iversen, L. L. & Emson, P. C. *Neuroscience* **6**, 1923–1934 (1981).
11. Jancso, G. & Kiraly, E. *J. comp. Neurol.* **190**, 781–792 (1980).
12. Nagy, J. I. *et al. Brain Res.* **186**, 435–444 (1980).
13. Gamse, R., Holzer, P. & Lembeck, F. *Br. J. Pharmac.* **68**, 207–213 (1980).
14. Helke, C. H., Dimico, J. A., Jacobowitz, D. M. & Kopin, I. J. *Brain Res.* **222**, 428–431 (1981).
15. Gamse, R., Leeman, S. E., Holzer, P. & Lembeck, F. N. S. *Archs Pharmac.* **317**, 140–148 (1981).
16. Jancso, G. & Jancso-Gabor, A. N. S. *Archs Pharmac.* **311**, 285–288 (1980).
17. Gibson, S. J., Polak, J. M., Wall, P. D. & Bloom, S. R. *J. comp. Neurol.* **201**, 65–79 (1981).
18. Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17**, 205–242 (1970).
19. Van der Loos, H. & Woolsey, T. A. *Science* **179**, 395–398 (1973).
20. Van der Loos, H. & Dörfel, J. *Neurosci. Lett.* **7**, 23–30 (1978).
21. Welker, C. *J. comp. Neurol.* **166**, 173–189 (1976).
22. Simons, D. J. *J. Neurophysiol.* **41**, 798–820 (1978).
23. Fitzgerald, M. *Brain Res.* (in the press).

Neural cell adhesion molecule is on embryonic muscle cells and mediates adhesion to nerve cells *in vitro*

Martin Grumet, Urs Rutishauser
& Gerald M. Edelman

Department of Molecular and Developmental Biology,
The Rockefeller University, 1230 York Avenue, New York,
New York 10021, USA

The interaction of nerves with muscles, particularly at neuromuscular junctions, is one of the most extensively studied areas of neurophysiology. From the point of view of developmental biology, however, little is known about the cellular and molecular events that lead to formation of these junctions^{1,2}. In previous investigations of cell–cell adhesion in the chick embryo, we identified a molecule on the surface of essentially all nerve cells, called neural cell adhesion molecule (N-CAM)^{3,4}, that appears to be a ligand in the formation of cell–cell bonds⁵. N-CAM is involved in a variety of developmental processes including neurite fasciculation and the formation of plexiform and cell layers in retinal tissue^{6–8}. We report here that N-CAM antigenic determinants are also present on chick embryo skeletal muscle cells. This observation raised the possibility that N-CAM is involved in adhesion between nerve and muscle cells. In model systems set up to explore this possibility, we observed that myoblasts will adhere *in vitro* to nerve cells from retina, and that this adhesion can be inhibited by F(ab') fragments of antibodies against N-CAM. In addition, membrane vesicles obtained from chick embryo brain or artificial vesicles⁵ reconstituted from lipid and affinity-purified N-CAM⁹ bound to the surface of myoblasts and myotubes; this binding was also blocked by anti-N-CAM F(ab'). These results suggest that the N-CAM adhesion system can facilitate nerve–muscle as well as nerve–nerve interactions *in vitro*.

Immunocytological studies carried out to determine the distribution of N-CAM in chick embryos revealed that the molecule is associated with nerve tissues; however, light staining was also observed in muscle⁶. As shown in Fig. 1, rabbit anti-N-CAM antibodies that had been affinity-purified on N-CAM–Sepharose⁵, bound to the surface membrane of cultured skeletal myoblasts and myotubes but not to fibroblasts in the same

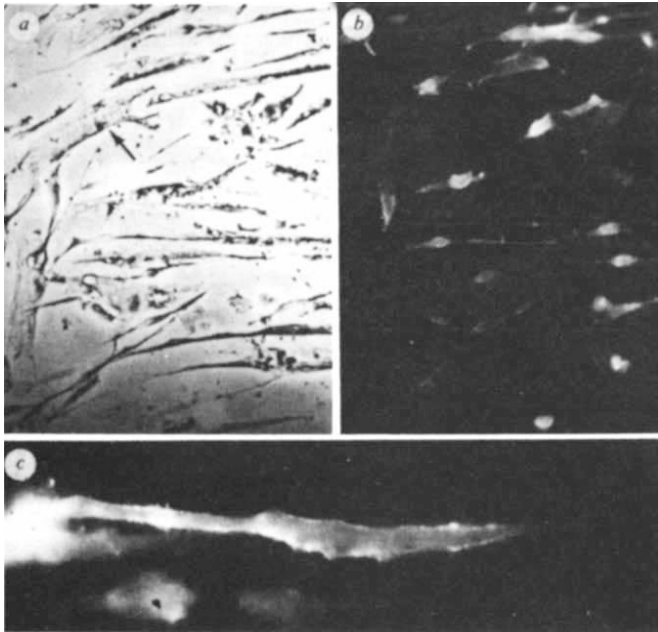


Fig. 1 Immunofluorescence staining of skeletal muscle cells with antibodies to N-CAM. *a*, Phase-contrast micrograph of myoblasts and early myotubes ($\times 264$), and *b*, fluorescence micrograph of the same field. *c*, Fluorescence micrograph of a myotube. $\times 880$. A fibroblastic cell which is not stained is indicated with an arrow. Muscle cells were obtained from 11–12-day White Leghorn chick embryo legs by trypsinization (0.1%, 30 min, 37°C) and trituration, and partially separated from fibroblasts by differential adhesion to culture dishes; the nonadherent cells were cultured on collagen-coated substrates for 2 days in Eagle's modified essential medium containing 10% horse serum and 5% chick embryo extract. For staining, the cultures were treated sequentially with 3.7% formaldehyde, 0.1 M glycine, 4% goat serum, affinity-purified rabbit anti-N-CAM IgG and fluorescein-labelled goat anti-rabbit IgG (Miles-Yeda). All reagents were prepared in phosphate-buffered saline; each antibody treatment was followed by four washes with buffer.

cultures. The staining of the muscle cells, which had been prefixed with formaldehyde, was mostly diffuse and lower in intensity than that seen previously on neurones⁶. Immunoglobulin from unimmunized rabbits did not give detectable staining.

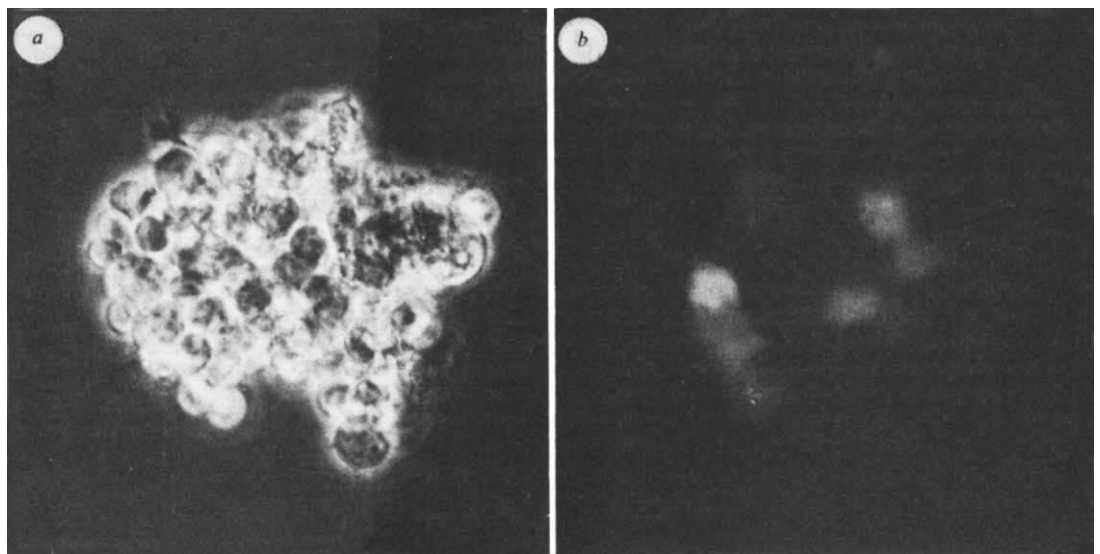
The presence of N-CAM antigenic determinants on muscle cells suggested that at some stage of their development they

might adhere to nerve cells through this molecule. This possibility was examined by studies of the adhesion of skeletal muscle myoblasts in suspension to monolayers¹⁰ of retinal nerve cells, of the co-aggregation of these cells in rotation–suspension cultures and of the binding of membrane vesicles containing N-CAM to myoblasts in suspension and to myotube monolayers. Although these assays represent highly simplified model systems for the analysis of cell–cell interactions in developing tissues, they have been useful in revealing and quantifying specific adhesion events, in particular for studying their molecular basis^{3,10,11}. The use of retina and brain for these studies reflects the requirement for substantial numbers of cells in the binding assays. Although retinal cells can establish a weak and transient synaptic connectivity with muscle *in vitro*¹², this system is of limited use in evaluating nerve–muscle interactions *in vivo*. Nevertheless, spinal cord cells are known to express N-CAM⁶, and the results obtained here will provide the necessary background to interpret more complex studies, such as the effects of anti-N-CAM antibody on cultures of muscle cells with spinal cord explants.

Adhesion between nerve cells has been measured previously using a cell suspension–cell monolayer binding assay¹⁰. In the present studies, myoblasts in suspension also bound to a retinal nerve cell monolayer, the level of binding during a 20-min incubation being about one-half that observed between nerve cells (Table 1). This binding was inhibited by anti-N-CAM F(ab') and did not require the presence of Ca^{2+} ; by these criteria it is distinguishable from the calcium-dependent mechanism of nerve cell adhesion¹³. Fibroblasts, which lack N-CAM⁶, did not bind in significant numbers to the retinal nerve cell monolayer.

Further evidence for nerve–muscle adhesion was obtained by incubating retinal and muscle cells together in suspension at 70 r.p.m. and 37°C for 20 min; in these conditions large aggregates were formed that contained both cell types (Fig. 2). Similarly, fluorescein-labelled membrane vesicles obtained from brain tissue bound to myoblasts and myotubes (Table 2, Fig. 3), although less strongly than to retinal neurones (Table 2). In each case, the interaction was inhibited by anti-N-CAM F(ab'); fibroblasts were included as a control and did not bind the vesicles. Table 2 and Fig. 3 show the binding to muscle cells of vesicles reconstituted from brain lipids and purified N-CAM. This adhesion was almost completely inhibited by anti-N-CAM antibody; vesicles containing lipids alone bound only in small numbers to a few cells. Binding was also observed between the N-CAM vesicles and cells obtained from heart muscle of 9-day embryos by trypsinization as described for the retinal cells. This suggests that N-CAM-mediated interactions may take place in a variety of muscle tissues of different origins.

Fig. 2 Co-aggregation of retinal nerve cells with fluorescein-labelled myoblasts. *a*, Phase-contrast micrograph, and *b*, fluorescence micrograph of the same field, $\times 840$. Cells (2.5×10^6 of each type) were incubated in 1 ml Eagle's minimal essential medium with spinner salts (SME medium) for 20 min at 37°C with rotary shaking at 70 r.p.m. Myoblasts were prepared as described in Fig. 1 legend, cultured in suspension for 4 h and further fractionated by density gradient centrifugation in Ficoll to yield a cell population consisting of about two-thirds myoblasts, the remaining cells probably being fibroblasts¹⁴. Myoblasts were internally labelled with diacetyl fluorescein¹⁰. Nerve cells were obtained from retina of 10-day chick embryos by mild trypsinization (0.002%, 15 min, 37°C)¹³.



We have shown previously⁵ that the adhesive specificity and the properties of the reconstituted N-CAM vesicles closely resemble those of intact nerve cells and have obtained evidence suggesting that cell-cell binding may involve interactions between N-CAM molecules on different cells. The fact that both muscle and nerve cells have N-CAM on their surfaces is consistent with this hypothesis. It also raises the possibility that muscle cells could bind to each other in part by the N-CAM mechanism. In preliminary studies, however, anti-N-CAM F(ab') reduced the binding of myoblasts to myoblasts or myotubes by only ~20%, which may indicate that other adhesion mechanisms and molecules are also involved, including those leading to the muscle-specific fusion of plasma membranes to form myotubes.

We conclude that N-CAM (or a similar molecule) is present on the plasma membrane of developing skeletal muscle cells, and that nerve cells can bind to muscle cells via the calcium-independent¹³ adhesion mechanism involving N-CAM. The presence of N-CAM on a cell type of mesenchymal origin as well

Table 1 Binding of neural and muscle cells to retinal nerve cell monolayers

Cell binding to monolayer	F(ab') present during binding*	No. of cells bound to monolayer (1 mm ²)	% Inhibition by anti-N-CAM F(ab')
Myoblasts	Nonimmune	174	—
Myoblasts	Anti-N-CAM	64	63
Retinal neurones	Nonimmune	389	—
Retinal neurones	Anti-N-CAM	105	73
Fibroblasts	—	30	—

Myoblasts and retinal nerve cells were prepared as described in Fig. 2 legend. Fibroblasts were dissociated from skin of 11-day chick embryos with 0.1% trypsin and grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum. At confluence, fibroblasts were removed from culture dishes with trypsin, cultured in suspension for 4 h and internally labelled with diacetyl fluorescein. As described previously¹⁰, monolayers of retinal cells were prepared on waxbean agglutinin-coated 35-mm culture dishes, incubated for 20 min at 25 °C with a suspension of cells internally labelled with fluorescein and scored for binding by fluorescence microscopy. Retinal nerve cells in suspension also bound to monolayers of myoblasts or myotubes, but only ~20% of this binding was inhibited by anti-N-CAM F(ab'). We have previously observed that inhibition of heterotypic binding by F(ab') is sometimes affected when the cell types in the suspension and monolayer are reversed¹³. Although the cause of this effect is not understood, binding of vesicles to myotubes in monolayers or to myoblasts in suspension was strongly inhibited by F(ab') (Table 2).

*20 µg ml⁻¹ of affinity-purified F(ab') or nonimmune F(ab') were added to the incubation medium.

Table 2 Binding of vesicles containing N-CAM to cells

Cells	Source of vesicles	% Of cells binding vesicles With	
		nonimmune F(ab')	With anti-N-CAM F(ab')
Myoblasts*	Brain tissue	69	18
Retinal neurones	Brain tissue	75	26
Fibroblasts	Brain tissue	0	—
Myoblasts	N-CAM + lipid	60	5
Myoblasts	Lipid	6	—

Myoblasts and retinal nerve cells were prepared as for Fig. 2, and fibroblasts as for Table 1, except without the internal fluorescent label. Vesicles containing N-CAM and carboxyfluorescein were isolated as described in Fig. 3 legend, and lipid vesicles were prepared as described but in the absence of N-CAM. Binding was scored by fluorescence microscopy (see Fig. 3 legend) after incubating 20 µl of vesicles with 5 × 10⁶ cells in 1 ml SME medium for 20 min at 37 °C in the presence of 10 µg ml⁻¹ of F(ab').

*60–70% of cells scored were myoblasts¹⁴.

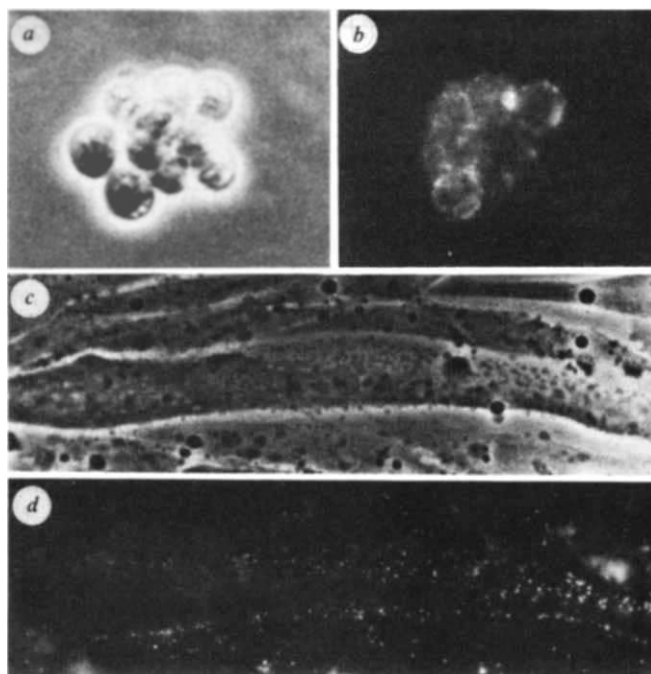


Fig. 3 Binding of brain tissue vesicles to myoblasts (a, b), and binding of reconstituted vesicles containing lipid and affinity-purified N-CAM to myotubes (c, d). a and c are phase contrast micrographs at ×570 and ×375 respectively; b and d are fluorescence micrographs of the same fields. Myoblasts were prepared as described in Fig. 2 legend, and myotubes as described in Fig. 1 legend. Fluorescent N-CAM vesicles were prepared by removal of detergent from a solution containing 6 mg ml⁻¹ of lipids extracted from chick brain, 0.2 mg ml⁻¹ of affinity-purified N-CAM and 5 mg ml⁻¹ 6-carboxyfluorescein. Membrane vesicles were prepared from 14-day embryo brains by homogenization of the tissue in the presence of 5 mg ml⁻¹ 6-carboxyfluorescein, isolated by sucrose density centrifugation and washed with phosphate-buffered saline. Binding of vesicles to cells was carried out by incubation of 20 µl of vesicles with 5 × 10⁶ cells in SME medium in a 1-ml suspension, or with myotubes on coverslips for 20 min at 37 °C with gentle shaking. Unbound vesicles were removed by differential centrifugation of the myoblasts or washing of myotube monolayers.

as in ectodermally derived nerve tissue raises the possibility that N-CAM expression on cells is more a reflection of functions required for morphogenesis than of cell lineage. In view of this possibility, it will be of particular interest to test for the presence and localization of N-CAM in embryos at earlier stages of development.

We thank Ms Helvi Hjelt for technical assistance. This work was supported by NIH grants (HD-09635, AI-11378, AM-04256) and a post-doctoral fellowship to M.G. (AI-06414).

Received 5 October 1981; accepted 11 January 1982.

- Nelson, P. G. *Physiol. Rev.* **55**, 1–61 (1975).
- Landmesser, L. T. A. *Rev. Neurosci.* **3**, 279–302 (1980).
- Brackenbury, R., Thiery, J.-P., Rutishauser, U. & Edelman, G. M. *J. biol. Chem.* **252**, 6835–6840 (1977).
- Thiery, J.-P., Brackenbury, R., Rutishauser, U. & Edelman, G. M. *J. biol. Chem.* **252**, 6841–6845 (1977).
- Rutishauser, U., Hoffman, S. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 685–689 (1982).
- Rutishauser, U., Thiery, J.-P., Brackenbury, R. & Edelman, G. M. *J. Cell Biol.* **79**, 371–381 (1978).
- Rutishauser, U., Gail, W. E. & Edelman, G. M. *J. Cell Biol.* **79**, 382–393 (1978).
- Buskirk, D. R., Thiery, J.-P., Brackenbury, R. & Edelman, G. M. *Nature* **285**, 488–489 (1980).
- Hoffman, S. *et al. J. biol. Chem.* (submitted).
- Rutishauser, U., Thiery, J.-P., Brackenbury, R., Sela, B.-A. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 577–581 (1976).
- Frazier, W. & Glaser, L. A. *Rev. Biochem.* **48**, 491–523 (1979).
- Paro, D. G., De Mello, F. G. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4977–4981 (1977).
- Brackenbury, R., Rutishauser, U. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 387–391 (1981).
- Turner, D. C. *Differentiation* **10**, 81–93 (1978).

Non-dopaminergic fibres may regulate dopamine-sensitive adenylate cyclase in the prefrontal cortex and nucleus accumbens

J. P. Tassin*, H. Simon†, D. Hervé*, G. Blanc*, M. Le Moal†, J. Glowinski* & J. Bockaert‡

* Groupe NB, INSERM U 114, and ‡ Laboratoire de Physiologie cellulaire, Collège de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France

† Laboratoire de Neurobiologie des comportements, Université de Bordeaux II, 33076 Bordeaux, France

The dopaminergic (DA) innervation of the prefrontal cortex in rat originates from neurones distributed in the anterior medial part of the ventral tegmental area (VTA)¹. Biochemical^{2,3} and electrophysiological⁴ studies have demonstrated that these mesocortico-prefrontal neurones are distinct from others located in the VTA which project to subcortical structures such as the nucleus accumbens. DA receptors coupled to an adenylate cyclase (D-1 receptors) are present in the prefrontal cortex and their topographical distribution closely resembles that of DA nerve terminals⁵⁻⁷. We have now further analysed the properties of the cortical D-1 receptors and compared them with those of a subcortical limbic structure examining, in the rat, the effects on DA-sensitive adenylate cyclase activity in the prefrontal cortex and the nucleus accumbens of (1) bilateral 6-hydroxydopamine (6-OHDA) microinjections in the VTA and (2) bilateral electrolytic lesions of the VTA. The results suggest that regulation of D-1 receptors is not solely dependent on the presynaptic DA innervation and that other non-DA fibres may contribute to it.

Male Sprague-Dawley rats (180–250 g) were lesioned bilaterally in the VTA either by 6-OHDA (4 µg in 1 µl) microinjections or by electrocoagulation as described previously^{8,9}. In some cases, animals were pretreated with desipramine (DMI; 25 mg per kg intraperitoneally, i.p.) 30 min before the 6-OHDA microinjections. Sham-operated rats were used as controls. All animals were killed between 11 a.m. and 1 p.m., 40–60 days after the lesions. Brains were rapidly removed, their rostral halves fixed with sodium chloride (0.9%) on a Leitz Kryomat¹⁰ and caudal parts kept for histological control. Using cylindrical tubes (1.4 and 0.9 mm diameter for the prefrontal cortex and the nucleus accumbens respectively), microdisks of tissues were punched out bilaterally on successive 500-µm-thick coronal slices from the ventral part of the prefrontal cortex⁷ (planes 11,300 and 10,800 µm according to the atlas of König and Klippel¹¹; four microdisks; 310 µg protein) or the nucleus accumbens (planes 9,600 and 9,100 µm; six microdisks; 190 µg protein). Tissues from individual animals were then treated to estimate DA-sensitive adenylate cyclase activity as previously described (refs 6, 7; Fig. 2 legend), and DA and noradrenaline levels in each sample were determined radioenzymatically¹².

As expected, a marked decrease in DA levels was observed in both the prefrontal cortex and the nucleus accumbens in rats with bilateral 6-OHDA lesions (Fig. 1). However, surprisingly, there was no change in the DA-sensitive adenylate cyclase activity in these structures despite the nearly complete destruction of their DA innervation (Fig. 1).

Similarly, in most of the animals with bilateral electrolytic lesions (10 out of the 14 lesioned rats) there was a significant reduction of DA levels in both the prefrontal cortex and the nucleus accumbens. In these rats, the DA-sensitive adenylate cyclase activity was significantly increased in the prefrontal cortex (+62 ± 6%; Fig. 1) when compared with results obtained in corresponding sham-operated animals. This effect resulted

from an increase in the maximal DA stimulation and not from a change in the apparent affinity of DA for its receptor, suggesting that the number of D-1 receptors coupled to the adenylate cyclase was enhanced (Fig. 2). In contrast, at a DA concentration (10^{-4} M) which induced the effect in the prefrontal cortex, no significant change in the DA-sensitive adenylate cyclase activity was observed in the nucleus accumbens (Fig. 1). This lack of effect was even seen in three animals in which the electrolytic lesions reduced DA levels by >95% (data not shown). These results seem to contradict those of Rosenfeld *et al.*¹³ who reported that unilateral electrolytic lesion of the VTA increases the DA-sensitive adenylate cyclase activity in the ipsilateral nucleus accumbens. However, they did not take into account the presence of high levels of DA in nucleus accumbens homogenates from the non-lesioned side. Indeed, a Lineweaver-Burk plot of their data reveals a difference in the affinity but no change in the maximal stimulation of the DA-sensitive adenylate cyclase activity in the denervated nucleus accumbens.

The difference in the effects of the 6-OHDA and electrolytic lesions on cortical noradrenergic (NA) levels might help to explain the lack of development of supersensitivity of the cortical D-1 receptors in the 6-OHDA lesioned rats. Indeed, in most cases, electrolytic VTA lesions did not destroy the ascending NA fibres projecting to the prefrontal cortex (Table 1). On the contrary, probably due to the neurotoxin diffusion, 6-OHDA microinjections also severely affected all ascending fibres of the NA dorsal bundle which pass near the VTA (Table 1). In another experiment designed to test further the eventual

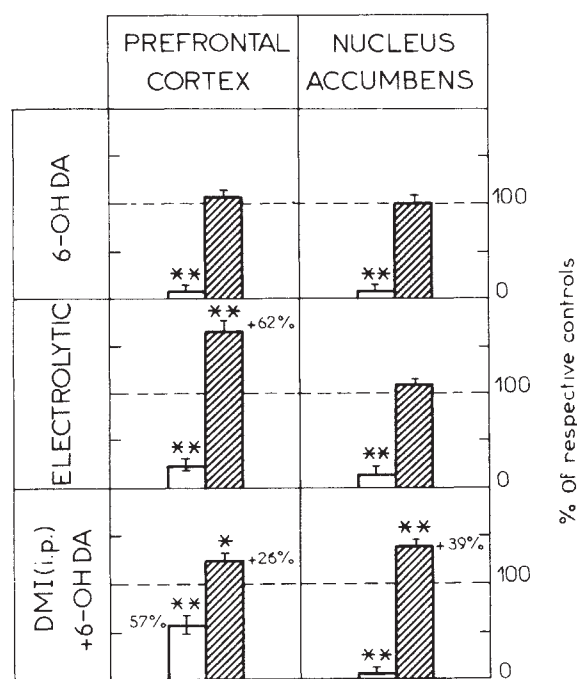


Fig. 1 Effects of different VTA lesions on DA levels and DA-sensitive adenylate cyclase activity in the prefrontal cortex and the nucleus accumbens. Rats were killed 40–60 days after each type of lesion and microdisks of tissue were punched out from the prefrontal cortex and the nucleus accumbens. After homogenization aliquots were taken for the estimation of DA levels and DA-sensitive adenylate cyclase activity. Results are the mean ± s.e.m. of data obtained with groups of 5–10 rats. Mean values ± s.e.m. of sham-operated animals were, respectively, in the prefrontal cortex and the nucleus accumbens: DA levels, 1.2 ± 0.1 and 75 ± 6 ng per mg protein; basal adenylate cyclase activities, 32 ± 1 and 38 ± 2 pmoles cyclic AMP per min per mg protein; DA (10^{-4} M)-stimulated adenylate cyclase activity, 63 ± 3 and 151 ± 12 pmoles cyclic AMP per min per mg protein. The basal adenylate cyclase activity in the lesioned animals was not affected in either the prefrontal cortex or the nucleus accumbens. * $P < 0.05$ and ** $P < 0.01$ (Student's *t*-test) when compared with respective values obtained in sham-operated animals. □, DA levels; ▨, DA-sensitive adenylate cyclase activity.

Table 1 Effects of different VTA lesions on NA levels in the prefrontal cortex

VTA lesion	Noradrenaline (ng per mg protein)	% Change compared with control
Sham	2.80 ± 0.20 (10)	
6-OHDA	0.10 ± 0.05 (10)	-96*
Electrolytic	2.29 ± 0.25 (10)	-18
DMI (i.p.) + 6-OHDA	1.85 ± 0.19 (8)	-34†

Forty to sixty days after each type of VTA lesion, rats were killed and, after sampling and homogenization of the prefrontal cortex tissues, NA levels were estimated radioenzymatically¹². Numbers in parentheses correspond to the size of each group. Values are mean ± s.e.m.

* $P < 0.001$ and † $P < 0.01$ when compared with sham-operated animals.

influence of the NA fibres on the regulation of cortical D-1 receptors, animals were treated with DMI before the microinjections of 6-OHDA in the VTA. As expected, this treatment markedly reduced the 6-OHDA neurotoxic effect on NA fibres but allowed the complete destruction of DA neurones projecting to the nucleus accumbens. Curiously, for an unknown reason, the cortical DA innervation was partially preserved because DA levels were only reduced by 43% in the prefrontal cortex. In these conditions, DA-sensitive adenylate cyclase activity was significantly increased in both the prefrontal cortex and the nucleus accumbens, the effect being more pronounced in the latter structure (Fig. 1).

A supersensitivity of D-1 receptors has already been shown in the striatum after electrolytically or 6-OHDA-induced degeneration of the nigro-striatal DA system^{14,15}. The results reported here—the marked increase in the DA-sensitive adenylate cyclase activity in the prefrontal cortex after electrolytic destruction of the mesocortico-prefrontal DA neurones—extend these observations. As already discussed, this effect probably results from an enhanced number of D-1 receptors. Its amplitude depends on the extent of DA denervation—in animals with electrolytic lesions a correlation was found between the increase in cortical DA-sensitive adenylate cyclase activity and the degree of the reduction in cortical DA levels (data not shown). It is thus not surprising that the increase in the cortical DA-sensitive adenylate cyclase activity was less pronounced after the partial destruction of the mesocortico-prefrontal DA neurones induced by the 6-OHDA lesion in DMI-pretreated rats. An interesting finding was that the presence of NA fibres seems necessary for the appearance of the increased cortical DA-sensitive adenylate cyclase activity; changes in the regulation of the cortical D-1 receptors were only seen when cortical NA fibres were only slightly or not affected by the lesion (electrolytic, 6-OHDA in DMI-pretreated rats). In contrast, in rats injected with 6-OHDA alone the concomitant destruction of the NA fibres prevented the increase in DA-sensitive adenylate cyclase activity despite the pronounced degeneration of the DA neurones innervating the prefrontal cortex. It is therefore possible that the cortical NA fibres exert a permissive role on the development of the cortical D-1 receptors supersensitivity. Of relevance in this context is the observation that there is a population of cells in the rat prefrontal cortex which responds to both DA and NA¹⁶.

Because NA accounts for ~2% of DA levels in the nucleus accumbens, it was not possible to estimate precisely the effect of 6-OHDA lesions on NA levels in this structure. However, changes in cortical NA levels can be used as an index of the destruction of ascending NA fibres which pass near or through the nucleus accumbens. In any case the development of the D-1 receptors' supersensitivity in the nucleus accumbens does not seem to depend on the presence of ascending NA fibres as there was no change in DA-sensitive adenylate cyclase activity in this structure either when NA fibres were not destroyed (electrolytic lesion) or when they were markedly affected (6-OHDA lesions). Although all types of lesion made induced a pronounced

degeneration of the nucleus accumbens DA innervation, the only one which resulted in a significant increase in the DA-sensitive adenylate cyclase activity in this structure was the 6-OHDA lesion made in DMI-pretreated rats. As this lesion was also the only one which partially prevented the destruction of the mesocortico-prefrontal DA neurones, we suggest that the presence of the mesocortico-prefrontal DA neurones is required for the development of D-1 receptors supersensitivity in the DA-denervated nucleus accumbens. Neurones originating in the prefrontal cortex are known to project into the nucleus accumbens¹⁷, and Pycocock *et al.*¹⁸ have recently shown that the specific destruction of DA terminals in the prefrontal cortex induced several biochemical changes in various subcortical structures, including an enhanced rate of DA utilization in the nucleus accumbens and the striatum. By their effect on cortical efferent pathways, the mesocortico-prefrontal DA neurones may contribute not only to the regulation of DA transmission in subcortical structures such as the nucleus accumbens, but also to the control of the activity of the target cells on which D-1 receptors are located. In other words, changes in the activity of cortical neurones projecting to the nucleus accumbens could be involved in the regulation of the DA-sensitive adenylate cyclase activity in this structure.

Several electrophysiological studies have demonstrated that the sensitivity of a receptor for a given transmitter can be modified by the action of another transmitter on the same cell^{19,20}. Moreover, sustained changes in cell responsiveness to non-monoaminergic transmitters have been shown after denervation of monoaminergic systems in the central nervous system²¹. Our biochemical results provide a novel example of a regulatory process which could intervene in the control of receptor sensitivity: we have shown that the expected increase in the DA-sensitive adenylate cyclase activity after DA denervation can be modulated or even prevented in situations in

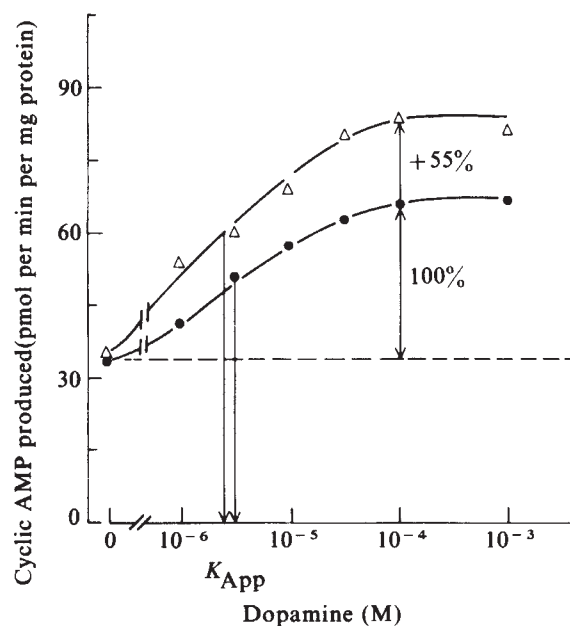


Fig. 2 Effect of bilateral electrolytic lesions of the VTA on the activity of the DA-sensitive adenylate cyclase in the prefrontal cortex. Eight microdisks of tissue (1.4 mm diameter) were punched out bilaterally in the ventral part of the frontal cortex⁷ (11,300 μm and 10,800 μm slices following the atlas of König and Klippel¹¹). Samples were homogenized in 200 μl of a 2 mM Tris-maleate, pH 7.2, 2 mM EGTA, 300 mM sucrose solution. Aliquots were taken to measure DA-sensitive adenylate cyclase activity^{6,7} and to estimate dopamine and noradrenaline levels¹². Each point corresponds to the mean of three determinations; extreme values did not differ by >5%. Variation of values obtained in 10 different sham-operated animals were within 8%. Δ , Electrolytic VTA lesion; $\bullet$, control.

which the target cells are simultaneously deprived or exposed to other significant signals. Such mechanisms may explain the variations observed^{14,15,22} in receptor supersensitivity, which may depend on the selectivity of the procedure used to destroy the afferent fibres. Although further evidence is required for such regulation of receptor sensitivity, the phenomenon may be functionally significant and should be taken into consideration when evaluating the efficacy of a given transmission.

This research was supported by grants from DRET (81050), INSERM (CRL 7953626) and Rhône Poulenc S.A.

Received 6 July 1981; accepted 12 January 1982.

1. Fuxe, K. *et al. Brain Res.* **82**, 349–355 (1974).
2. Hervé, D. *et al. Neurosci. Lett.* **15**, 127–133 (1979).
3. Lisoprawski, A., Hervé, D., Blanc, G., Glowinski, J. & Tassin, J. P. *Brain Res.* **183**, 229–234 (1980).
4. Deniau, J. M., Thierry, A. M. & Féger, J. *Brain Res.* **189**, 315–326 (1980).
5. Von Hungen, K. & Roberts, S. *Eur. J. Biochem.* **36**, 391–401 (1973).
6. Bockaert, J., Tassin, J. P., Thierry, A. M., Glowinski, J. & Prémont, J. *Brain Res.* **122**, 71–86 (1977).
7. Tassin, J. P. *et al. Brain Res.* **154**, 241–251 (1978).
8. Le Moal, M., Cardo, B. & Stinus, L. *Physiol. Behav.* **4**, 567–573 (1969).
9. Galey, D., Simon, M. & Le Moal, M. *Brain Res.* **124**, 83–97 (1977).
10. Tassin, J. P. *et al. Brain Res.* **141**, 267–281 (1978).
11. König, J. H. R. & Klippel, R. A. *A Stereotaxic Atlas of the Forebrain and Lower Parts of the Brain Stem* (William and Wilkins, Baltimore, 1963).
12. Gauchy, C., Tassin, J. P., Glowinski, J. & Chéramy, A. *J. Neurochem.* **26**, 471–480 (1976).
13. Rosenfeld, M. R., Seeger, T. F., Sharpless, N. S., Gardner, E. L. & Makman, M. H. *Brain Res.* **173**, 572–576 (1979).
14. Mishra, R. K., Gardner, E. L., Katzman, R. & Makman, M. H. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3883–3887 (1974).
15. Prémont, J., Tassin, J. P., Thierry, A. M. & Glowinski, J. *Expl Brain Res.* **23**, Suppl. 165 (1975).
16. Bunney, B. S. & Aghajanian, G. K. *Life Sci.* **19**, 1783–1792 (1976).
17. Becksteadt, R. M. *J. comp. Neurol.* **184**, 42–43 (1979).
18. Pycok, C. J., Carter, C. J. & Kerwin, R. W. *J. Neurochem.* **34**, 91–99 (1980).
19. Ferron, A., Descaries, L. & Reader, T. A. *Brain Res.* **231**, 93–108 (1982).
20. Moises, H. C., Woodward, D. J., Hoffer, B. J. & Freedman, T. *Expl. Neurol.* **64**, 493–515 (1979).
21. Wang, R. Y., De Montigny, C., Gold, B. I., Roth, R. H. & Aghajanian, G. K. *Brain Res.* **178**, 479–497 (1979).
22. Krueger, B. K., Forn, J., Walters, J. R., Roth, R. H. & Greengard, P. *Molec. Pharmac.* **12**, 639–648 (1976).

Importance of amino acid uptake and decarboxylation in gastrin release from isolated G cells

Lenard M. Lichtenberger, Remi Delansorne & Lynne A. Graziani

Department of Physiology and Cell Biology, University of Texas Medical School at Houston, Houston, Texas 77025, USA

Studies of various species have demonstrated that the post-prandial rise in serum gastrin levels is primarily attributable to the proteinaceous substances ingested^{1–5}. Although it is generally assumed that dietary amino acids are the primary stimulants of gastrin release, evidence supporting this notion based on experiments *in vivo* has been conflicting and unconvincing^{6–10}. We have investigated the role of amino acids as stimulants of gastrin release using a preparation of isolated rodent gastrin-containing G cells^{11,12}. We report here that the ability of an amino acid to stimulate gastrin release *in vitro* is directly correlated with its lipid solubility, the aromatic and long chain aliphatic amino acids being the most potent. In addition, we provide evidence that amino acid decarboxylation may be a necessary step in gastrin release as inhibitors of decarboxylase activity abolished amino acid-induced gastrin release, and amines, the decarboxylated derivatives of amino acids, were potent stimulants of hormone secretion *in vitro*. Thus, the uptake and decarboxylation of amine precursors may be a functionally important property of cells belonging to the APUD (amine precursor uptake and decarboxylation) class of endocrine cells.

We investigated the effects of amino acids and their metabolites on gastrin release *in vitro* using a preparation of enriched G cells isolated according to a modification of a procedure described elsewhere^{11,12} (see Fig. 1 legend). Initially we compared the ability of various amino acids to stimulate gastrin release *in vitro*. We included one or more of the amino acids from each of the major classes. The results (see Table 1) indicate that both charged amino acid residues and simple neutral aliphatic amino acid residues having a 1–3 carbon backbone failed to induce gastrin release from isolated G cells. In contrast, the more nonpolar molecules, the aliphatic amino acids having four or more carbon atoms, and the aromatic amino acids, stimulated gastrin release into the medium.

The effectiveness of an amino acid in stimulating gastrin release correlated significantly ($r < 0.001$) with the relative hydrophobicity of its side chain (Fig. 1). This relationship between the lipid solubility of an amino acid and its efficacy in stimulating gastrin release *in vitro* suggests that hormone secretion may depend on the ability of the amino acid to diffuse from the lumen into the interior of the G cell. This notion is supported by reports that amino acids and related compounds are transported into the gastric epithelium by a passive transport process¹³. The intracellular mechanism by which amino acid uptake is coupled to gastrin secretion, however, has yet to be characterized.

Table 1 Effect of amino acids and amines on gastrin release *in vitro*

Amino acids	Gastrin release (% basal)	Amines	Gastrin release (% of basal)
$\begin{array}{c} \text{R} \\ \\ \text{NH}_2-\text{CH}-\text{COOH} \end{array}$		$\begin{array}{c} \text{R} \\ \\ \text{NH}_2-\text{CH}_2 \end{array}$	
<i>Aliphatic</i>			
Glycine	98 ± 22(4)	Methylamine	350 ± 77(5)*†
Alanine	76 ± 20(4)	Ethylamine	455 ± 98(5)*†
Valine	101 ± 11(10)	Isobutylamine	306 ± 30(10)*†
Leucine	134 ± 5(4)*		
Isoleucine	176 ± 49(4)		
<i>Hydroxy</i>			
Serine	87 ± 6(5)		
<i>Aromatic</i>			
Phenylalanine	186 ± 26(5)*	Methylbenzylamine	252 ± 46(5)*
Tryptophan	212 ± 23(5)*		
<i>Charged</i>			
Arginine	99 ± 27(4)		
Ornithine	65 ± 15(5)		
Glutamine	109 ± 6(4)		
Cysteine	89 ± 10(9)	Cysteamine	200 ± 25(10)*†
Glutamic Acid	87 ± 7.3(4)	δ -Aminobutyric acid	126 ± 24(5)
<i>Imino</i>			
Proline	143 ± 24(4)		

An isolated and enriched preparation of rodent antral G cells was prepared using a modification of a procedure described elsewhere^{11,12}. Strips of antral tissue were incubated for 90 min in enriched KRB (Krebs-Ringer bicarbonate) buffer containing pronase E (0.125%; EM Laboratories), with shaking in water bath at 37 °C in an atmosphere of 95% O₂/5% CO₂ at 105 cycles min⁻¹. At the end of this period the remaining tissue was vigorously shaken by hand (~2,000 cycles min⁻¹) in pronase-containing medium, and then lightly scraped with a glass slide. This combination of selective pronase digestion followed by high shearing forces results in the removal of cells from the basal elements of the antral gland. The exfoliated cells and tissue were then incubated in enriched KRB buffer containing 20 µg ml⁻¹ of DNase I for 30 min, followed by gentle filtration (70 µm pore size) to disrupt the cell aggregates. The resulting cell suspension was washed several times in fresh enzyme-free KRB, and counted using a haemocytometer. Gastrin secretory testing was initiated by injecting the cells (50–150 × 10³ per tube) into a siliconized tube containing amino acid-free KRB with and without (control) the test agent. All test agents were used at a final concentration of 10 mM. The cells were incubated with shaking in a water bath at 37 °C in an atmosphere of 95% O₂/5% CO₂ for 10 min. The cells and medium were separated by filtration, and the filtrate was heat extracted at 96 °C for 10 min (ref. 2). The medium samples were then stored at –20 °C before radioimmunoassay of the medium gastrin concentration. Gastrin concentration was determined by a modification of the technique of Yalow and Berson¹⁹. Purified human gastrin G-17 I was iodinated by the chloramine T method and purified by the method of Stadil and Rehfeld²⁰. Antibody 1296, which recognizes all known molecular forms of gastrin²¹, was used in all assays. Numbers in parentheses are number of tubes per group.

* Represents a statistically significant difference ($P < 0.05$) compared with medium gastrin concentration of cells incubated in control medium.

† Represents a statistically significant difference ($P < 0.05$) compared with medium gastrin concentration for cells incubated in medium containing the respective amino acid (carboxylated derivative of the amine).

Pearse initially proposed the APUD hypothesis¹⁴. He demonstrated that many endocrine cells, like neural tissue, are able to decarboxylate amine precursors to form the parent amine. Many endocrine cells of the gastrointestinal tract, including the G cell, have been demonstrated histochemically to be members of the APUD family of endocrine cells. Although the importance of the decarboxylase enzymes has been established neurophysiologically, their functional role in endocrine cells remains to be elucidated fully.

We have also demonstrated (see Table 1) that many amines, products of amino acid decarboxylation, are highly efficient in stimulating gastrin release *in vitro*. Note that without exception, amines induced a greater stimulation of hormone secretion than the respective amino acids from which they are derived. These findings are not totally unexpected as previous studies have demonstrated that certain aromatic and aliphatic amines (the catecholamines and cysteamine) are potent stimulants of gastrin release *in vivo*²³⁻²⁵. Subsequent experiments have indicated that certain biogenic amines, (serotonin, histamine, spermine and taurine) are unable to stimulate gastrin release *in vitro*. These results suggest that amine-stimulated gastrin release is not a nonspecific G cell response to amines in general, but is a specific response to certain molecular species of this class of compounds.

We then investigated whether the ability of an amino acid to stimulate gastrin release *in vitro* could be modulated in parallel with the decarboxylase activity of the endocrine cell. To test this possibility, glandular cells of the antrum were incubated in medium containing the stimulatory aromatic amino acid, phenylalanine, with or without the addition of the decarboxylase inhibitors, α -methyl-L-dopa (specific L-aromatic amino acid decarboxylase inhibitor), 4-deoxyripyridoxine and aminooxyacetic acid (nonspecific decarboxylase inhibitors)^{15,16}. Subsequent studies revealed that all three agents significantly reduced the L-aromatic amino acid decarboxylase activity of isolated intact antral glandular cells when [1-alanine-¹⁴C]DL-dopa was used as substrate (unpublished results). Figure 2 shows that all three inhibitors significantly blocked phenylalanine-induced gastrin release. In addition, this abolition of amino acid-induced hormone release could not be attributed to a general depression of cell function, as the decarboxylated derivative of phenylalanine, methylbenzylamine, was still capable of stimulating gastrin release in the presence of deoxyripyridoxine. Note also that subsequent experiments demonstrated that these inhibitors of cellular decarboxylase activity had minimal effect on basal gastrin release, and effectively

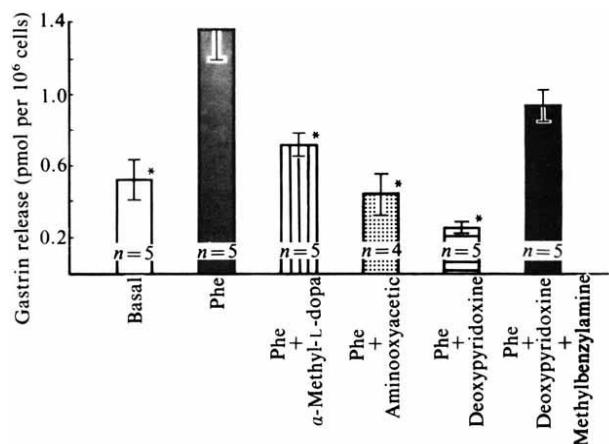


Fig. 2 Effect of decarboxylase inhibitors on phenylalanine-induced gastrin release *in vitro*. The isolated and enriched G cells were prepared and studied as described in Table 1 legend. In addition to cells incubated in amino acid-free KRB alone (basal), cells were incubated in phenylalanine-containing medium (10 mM) in the absence and presence of the following inhibitors of decarboxylase activity: α -methyl-L-DOPA (0.01 mM); aminooxyacetic acid (1 mM); 4-deoxyripyridoxine (1 mM) alone, and together with the stimulatory amine, methylbenzylamine (10 mM). * Represents a statistically significant difference ($P < 0.05$) compared with the medium gastrin levels of cells incubated with phenylalanine alone. † Represents a statistically significant difference ($P < 0.001$) compared with the medium gastrin levels of cells incubated with phenylalanine and deoxyripyridoxine.

abolished the stimulatory actions of other aromatic amino acids (such as tryptophan) on gastrin release *in vitro*.

These results suggest that gastrin release, which is triggered by the ingestion of a meal containing protein, may depend in part on the passive transport of amino acids into the G cell, followed by their intracellular decarboxylation. This notion is further supported by autoradiographic evidence^{17,18} that the DOPA decarboxylase activity of the G cell resides anatomically in close proximity to the gastrin-containing secretory granules. However, the mechanism by which the amine product is coupled to hormone release remains to be determined. Further experiments will ascertain whether the decarboxylase enzymes have a similar role in mediating the amino acid-induced release of hormones from other endocrine cells possessing the APUD facility.

We thank Mr William Vargo for technical assistance and Drs Morton I. Grossman and John H. Walsh (CURE) for donating purified human gastrin G-171 and antibody 1296 which were used in the radioimmunoassays. This work was supported by NIH grant AM 20686. R.D. performed this work as a visiting student from Ecole Nationale Veterinaire D'Alfort.

Received 31 August; accepted 23 December 1981.

1. Saint-Hilaire, S., Lavers, M. K., Kennedy, J. & Code, C. F. *Gastroenterology* **39**, 1-11 (1960).
2. Debas, H. T., Scendes, A., Walsh, J. H. & Grossman, M. I. in *Endocrinology of the Gut* (eds Chey, W. Y. & Brooks, S. P.) 222-232 (Charles B. Slack, New Jersey, 1974).
3. Debas, H. T., Walsh, J. H. & Grossman, M. I. in *Gastrointestinal Hormones* (ed. Thompson, J. C.) 425-435 (University of Texas Press, Texas 1975).
4. Richardson, C. T., Walsh, J. H., Hicks, J. I. & Fordtran, J. S. *J. clin. Invest.* **58**, 623-631 (1976).
5. Walsh, J. H. *Fedn. Proc.* **36**, 1948-1951 (1976).
6. Walsh, J. H. & Grossman, M. I. *N. Engl. J. Med.* **292**, 1377-1384 (1975).
7. Elwin, C. E. & Uvnas, B. in *Gastrin* (ed. Grossman, M. I.) 69-82 (UCLA Press, Los Angeles 1966).
8. Elwin, C. E. *Scand. J. Gastroent.* **9**, 239-247 (1974).
9. Konturek, S. J., Tasler, J., Cieszkowski, M., Dobrzanska, M. & Wunsch, E. *Am. J. Physiol.* **233**, 170-174 (1977).
10. Strunz, U. T., Walsh, J. H. & Grossman, M. I. *Proc. Soc. exp. Biol. Med.* **157**, 440-441 (1978).
11. Forssmann, W. G., Lichtenberger, L. M., Helmstaedter, V. & Ito, S. *Cell Tissue Res.* **200**, 163-177 (1979).
12. Lichtenberger, L. M., Forssmann, W. G. & Ito, S. *Gastroenterology* **79**, 447-459 (1980).
13. Schanker, L. S. *J. Medl pharm. Chem.* **2**, 343-359 (1960).
14. Pearse, A. G. E. *J. Histochem. Cytochem.* **17**, 303-313 (1969).
15. Holtz, P. & Palm, D. *Pharmac. Rev.* **16**, 113-178 (1964).

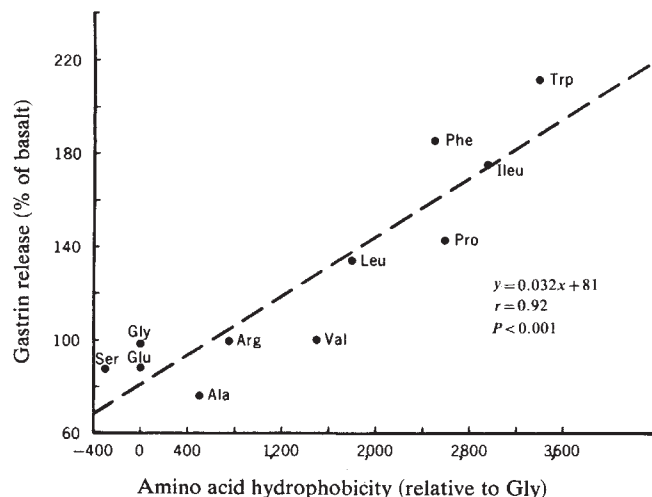


Fig. 1 Relationship between the ability of an amino acid to stimulate gastrin release *in vitro* and the hydrophobic properties of its side chain. The data points were limited to those amino acids whose relative hydrophobicity were estimated by Bigelow and Channon²². There is a statistically significant linear relationship between the two properties ($r = 0.92$; $P < 0.001$).

16. Lovenberg, W., Weisback, H. & Udenfriend, S. *J. biol. Chem.* **237**, 89–93 (1962).
17. Kobayashi, S. *Archs Histol. Jap.* **37**, 313 (1975).
18. Rubin, W. & Schwartz, B. *Gastroenterology* **81**, 311–320 (1981).
19. Yalow, R. S. & Berson, S. A. *Gastroenterology* **58**, 1–14 (1970).
20. Stadil, F. W. & Rehfeld, J. F. *Scand. J. clin. Invest.* **30**, 361–368 (1972).
21. Dockray, G. J. & Walsh, J. H. *Gastroenterology* **68**, 222–230 (1975).
22. Bigelow, C. C. & Channon, M. in *Handbook of Biochemistry and Molecular Biology* 3rd edn (ed. Fasman, G. D.) 209 (CRC, Ohio 1976).
23. Stadil, F. & Rehfeld, J. F. *Gastroenterology* **66**, 7–15 (1974).
24. Hayes, J. R., Ardill, J., Kennedy, L., Buchanan, K. D. & Shanks, R. G. *Lancet* **i**, 819–821 (1972).
25. Lichtenberger, L. M., Szabo, S. & Trier, J. S. *Gastroenterology* **73**, 1305–1308 (1977).

Inhibition of spindle elongation in permeabilized mitotic cells by erythro-9-[3-(2-hydroxynonyl)] adenine

W. Zacheus Cande

Department of Botany, University of California, Berkeley, California 94720, USA

Studies using living cells^{1,2} and experimental manipulation of permeabilized mitotic cells^{3,4} demonstrate that the movement of anaphase chromosomes consists of two physiologically distinct events: chromosome movement to the spindle poles (anaphase A) and separation of the poles via spindle elongation (anaphase B). Shortening of kinetochore-attached microtubules is associated with anaphase A and rearrangement and lengthening of the polar microtubules, especially those found in the spindle midzone, with anaphase B^{5,6}. It has been suggested that interactions between microtubules, mediated by dynein-like cross-bridges similar to those found in flagella, may be the mechanochemical system responsible for anaphases A and B^{7,8}. Recently Bouchard *et al.*⁹ have demonstrated that erythro-9-[3-(2-hydroxynonyl)]adenine (EHNA), a protein carboxyl-methylation inhibitor, is a specific inhibitor of flagellar beat and dynein ATPase activity. I have now investigated the possibility that dynein-like ATPases have a role in anaphase by adding EHNA to permeabilized cell models in conditions compatible with maintenance of anaphases A and B after lysis^{3,4}. I find that EHNA blocks anaphase B but not anaphase A. These results are consistent with the involvement of a dynein-like ATPase in spindle elongation.

Mitotic PtK₁ cells (a kidney epithelial line derived from the rat kangaroo) continue to undergo anaphase after permeabilization of the cell with the nonionic detergent Brij 58 (Fig. 1, step 1, and ref. 3) provided polyethylene glycol is added to the medium 1–2 min after lysis (Fig. 1, step 2). After lysis the dividing cells are permeable to ions, nucleotides, dyes and proteins^{3,4}. The rate and extent of spindle elongation (Fig. 1a, ●) in lysed mitotic cells depend on the ATP concentration in the medium⁴. In contrast, chromosome-to-pole movement in the permeabilized cell (Fig. 1a, △) does not require Mg²⁺-ATP and is not affected by the presence of sulphhydryl reagents, vanadate (a flagellar dynein ATPase inhibitor) and non-hydrolysable ATP analogues in the lysis medium⁴.

EHNA, at concentrations that block movement of native rat and sea urchin spermatozoa⁹, does not block anaphase when added to intact PtK₁ cells during prometaphase.

After lysis, EHNA preferentially blocks spindle elongation in dividing cells (Fig. 1b, Table 1) but does not reduce the rate of chromosome-to-pole movement. Inhibition of anaphase B by EHNA can be prevented by increasing the level of ATP in the lysis medium (Table 1). This result is consistent with the observation that EHNA is a weak mixed competitive inhibitor of dynein ATPase activity⁹. The relative concentration of EHNA required for 75% inhibition of anaphase B (0.2 mM ATP, 0.5 mM EHNA; Table 1) is similar to that required for 50% inhibition of dynein ATPase activity (0.1 mM ATP, 0.23 mM

Table 1 Effect of EHNA on chromosome movement

Treatment	No. of experiments	Rates of movement ($\mu\text{m min}^{-1}$)		
		Chromosome separation	Spindle elongation	Chromosome-to-pole movement
0.2 mM ATP	14	0.55 ± 0.15	0.26 ± 0.10	0.15 ± 0.07
+0.1 mM EHNA	6	0.55 ± 0.16	0.19 ± 0.14	0.18 ± 0.05
+0.5 mM EHNA	9	0.46 ± 0.15	0.06 ± 0.10	0.19 ± 0.06
+2.0 mM EHNA	13	0.32 ± 0.11	0.04 ± 0.06	0.14 ± 0.05
+2.0 mM EHNA and 1.25 mM ATP	5	0.59 ± 0.14	0.33 ± 0.06	0.14 ± 0.04

Rates of movement were estimated by measuring the slopes of the graphs (Fig. 1) of chromosome-chromosome and pole-pole distances from the moment of addition of the step 2 medium (see Fig. 1 legend) to the position of maximum displacement of chromosomes or poles. The rate of chromosome-to-pole movement was calculated as one-half the difference between the rates of chromosome separation and spindle elongation³. Values are given with standard deviations.

EHNA)⁹ and reactivation of demembrated sea urchin spermatozoa (0.2 mM ATP, ~0.75 mM EHNA)⁹.

EHNA was originally described as a potent inhibitor of protein carboxylmethylation because it increases the intracellular concentration of S-adenosyl-L-homocysteine through inhibition of adenosine deaminase^{9–11}. To study whether EHNA blocked anaphase by this mechanism, PtK₁ cells were lysed in S-adenosyl-L-homocysteine, adenosine and cycloleucine, which are known to inhibit protein carboxylmethylation^{9–11}, and in S-adenosyl-L-methionine, a donor of methyl groups¹¹. As S-adenosyl-L-homocysteine and adenosine are almost as effective as EHNA in retarding spindle elongation (Table 2), we cannot rule out the possibility that protein carboxylmethylation is somehow involved in this process. However, the concentration of these drugs required to inhibit anaphase B are 10²–10³-fold higher than that required to inhibit protein carboxylmethylation *in vivo* or *in vitro*^{9–11}. Cycloleucine, another inhibitor, does not block anaphase B (Table 2), and S-adenosyl-L-methionine, which should promote protein carboxylmethylation *in vitro*, inhibits anaphase B (Table 2) in the permeabilized cell. In addition, inhibition of spindle elongation by EHNA and adenosine depends on the ATP concentration in the lysis medium (Tables 1, 2). These results suggest that EHNA, S-adenosyl-L-homocysteine and adenosine all inhibit spindle elongation due to a common property other than their ability to inhibit protein carboxylmethylation. In this regard, spindle elongation differs from flagellar movement, because dynein ATPase activity and flagellar beat are not inhibited by adenosine⁹.

It has been suggested that actomyosin may help to generate the forces required for anaphase chromosome movement¹². As described previously¹³, skeletal myosin ATPase activity is unaffected by millimolar concentrations of EHNA, and glycerinated myofibril contraction continues in concentrations of ATP and EHNA that block anaphase B (W.Z.C., unpublished data). To study the effect of EHNA on a process in lysed cells known to involve actomyosin, 1–2 mM EHNA was added to permeabilized PtK₁ cells undergoing cytokinesis. Cleavage furrow constriction continues for several minutes after lysis using

Table 2 Effect of carboxyl transmethylation inhibitors on chromosome movement

Treatment	No. of experiments	Rates of movement ($\mu\text{m min}^{-1}$)		
		Chromosome separation	Spindle elongation	Chromosome-to-pole movement
0.2 mM ATP	5	0.56 ± 0.08	0.24 ± 0.09	0.16 ± 0.02
+0.5 mM S-adenosyl-L-homocysteine	5	0.37 ± 0.15	0.11 ± 0.16	0.13 ± 0.03
+0.5 mM S-adenosyl-L-methionine	5	0.43 ± 0.10	0.09 ± 0.08	0.18 ± 0.06
+12 mM cycloleucine	5	0.61 ± 0.10	0.21 ± 0.11	0.19 ± 0.06
+0.5 mM adenosine	5	0.39 ± 0.09	0.14 ± 0.16	0.13 ± 0.04
+2.0 mM adenosine and 1.25 mM ATP	3	0.62 ± 0.20	0.33 ± 0.12	0.15 ± 0.04

the procedure described in Fig. 1 legend (see ref. 3) and EHNA does not alter the rate or extent of cleavage (W.Z.C., unpublished data).

As microtubule polymerization may have a role in spindle elongation (reviewed in refs 2, 5), the effect of EHNA on *in vitro* polymerization of neurotubulin was followed spectrophotometrically by monitoring changes in turbidity¹⁴. Extracts containing 3–5 mg ml⁻¹ twice-cycled neurotubulin³ were warmed to 37°C in medium containing 100 mM PIPES, pH 6.94, 1 mM MgSO₄, 1 mM EGTA and either 0.5 mM GTP and 2 mM EHNA or 1 mM GTP and 5 mM EHNA. No change

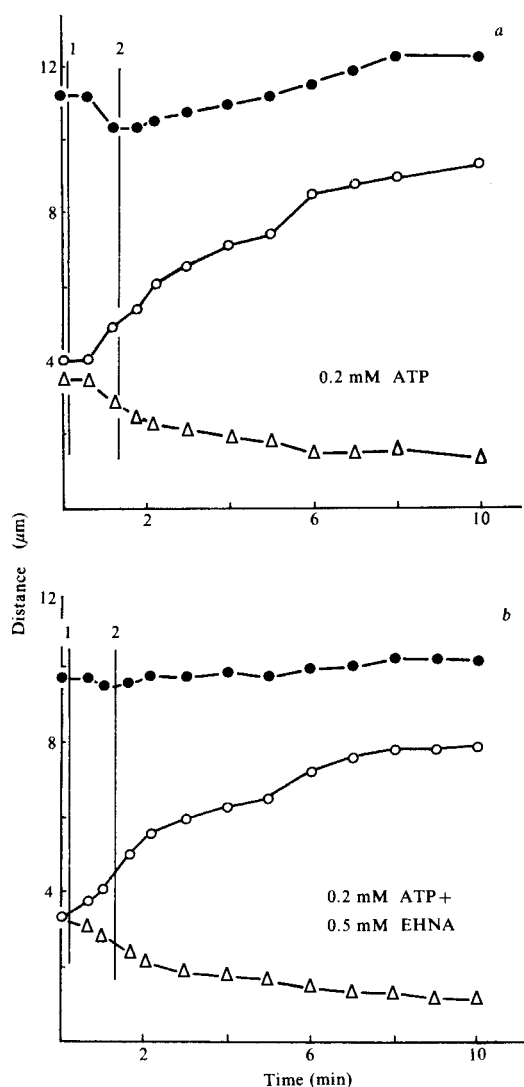


Fig. 1 Effect of EHNA on chromosome movements in permeabilized cells. Separation of spindle poles (●), sister chromatids (○) and chromosome-to-pole movement in one half spindle (△) when cells are lysed in a medium containing 0.2 mM ATP (a) or 0.2 mM ATP and 0.5 mM EHNA (b). PtK₁ cells were used in all experiments and were maintained and handled for light microscopy as described in ref. 3. Coverslips were mounted on slides with coverslip fragments as spacers. Anaphase cells were lysed by flushing solutions under the coverslip in a two-step lysis procedure: step 1 medium contained 85 mM PIPES, 0.08% Brij 58, equimillimolar nucleotide or EHNA (Burroughs-Wellcome) and MgSO₄, 10⁻⁷ M free Ca²⁺ and 1 mM free Mg²⁺. Step 2, which followed 70–90 s after step 1, used a similar medium that included in addition 2.5% polyethylene glycol (molecular weight 20,000) and 0.1% Brij 58. The composition of the Ca²⁺-EGTA buffer system was determined using a computer program provided by R. Steinhardt. Using a stability constant of 10.7 at pH 6.94 and 10 mM EGTA as a buffer, 1.51 mM CaCl₂ and 1.28 mM MgSO₄ were added to the lysis medium.

in rate or extent of microtubule polymerization in the presence of four- to fivefold more EHNA than nucleotide was detected.

Thus EHNA inhibition of anaphase B is consistent with the argument that a dynein-like ATPase is responsible for generating the forces required for spindle elongation^{7,8} and that anaphases A and B are physiologically distinct processes that can be uncoupled by the appropriate experimental manipulation^{1,3,4}.

I thank Susan Elliger for technical assistance, Carl Elliger and Kent McDonald for helpful discussions, and Steve Penningroth for preprints describing the specificity of EHNA inhibition of dynein ATPase activity and helpful discussions concerning the mechanism of EHNA inhibition. This research was supported by grants from NIH, American Cancer Society and the University of California Cancer Research Coordinating Committee.

Received 10 August; accepted 21 December 1981.

1. Ris, H. *Biol. Bull.* **96**, 90–106 (1949).
2. Inoué, S. & Ritter, H. in *Molecules and Cell Movement* (eds Inoué, S. & Stephens, R. E.) 3–30 (Raven, New York, 1975).
3. Cande, W. Z., McDonald, K. & Meeusen, R. L. *J. Cell Biol.* **88**, 618–629 (1981).
4. Cande, W. Z. *Cell* (in the press).
5. Fuge, H. *Int. Rev. Cytol.*, Suppl. **6**, 1–58 (1977).
6. McDonald, K., Pickett-Heaps, J. D., McIntosh, J. R. & Tippet, D. H. *J. Cell Biol.* **74**, 377–388 (1977).
7. McIntosh, J. R., Hepler, P. K. & van Wic, D. G. *Nature* **224**, 659–663 (1969).
8. Margolis, R. L., Wilson, L. & Kiefer, B. I. *Nature* **272**, 450–452 (1978).
9. Bouchard, P., Penningroth, S. M., Cheung, A., Gagnon, C. & Bardin, C. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1033–1036 (1981).
10. Berthelsen, S. & Pettinger, W. A. *Life Sci.* **21**, 595–606 (1977).
11. O'Dea, R. F., Viveros, O. H., Diliberto, E. J. Jr *Biochem. Pharmac.* **30**, 1163–1168 (1981).
12. Forer, A. in *Nuclear Division in the Fungi* (ed. Heath, I. B.) 21–68 (Academic, New York, 1978).
13. Penningroth, S. M., Cheung, A., Bouchard, P., Gagnon, C. & Bardin, C. W. *Biochem. biophys. Res. Commun.* (in the press).
14. Gaskins, F. & Cantor, C. R. *J. molec. Biol.* **89**, 737–758 (1974).

Inhibitors of dynein activity block intracellular transport in erythrophores

Mary C. Beckerle & Keith R. Porter

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Campus Box 347, Boulder, Colorado 80309, USA

Erythrophores, cells capable of translocating pigment granules along a radially arrayed cytoskeletal framework, provide an excellent model system for studying intracellular particle transport. Numerous examples of intracellular motility are based on microtubules¹ and all available evidence indicates that organized pigment granule transport in erythrophores depends on these structures². The best studied microtubule-based motility system is the flagellum. As it relies on a dynein-microtubule interaction to generate motive force³, we have investigated the effects of dynein ATPase inhibitors on pigment granule transport. Vanadate is a potent inhibitor of dynein ATPase activity *in vitro*^{4–6} and has been shown to inhibit microtubule-based motility in cilia⁴, flagella⁴ and the mitotic spindle⁷. Recently, another dynein inhibitor, erythro-9-3-(2-hydroxynonyl)adenine (EHNA), has been described⁸. We report here that both microinjected vanadate and exogenously applied EHNA block saltatory movement of pigment granules in isolated squirrelfish erythrophores. The vanadate and EHNA sensitivities of directed pigment motion suggest that a dynein-like molecule may be an essential component of the erythrophore intracellular motility system.

Erythrophores isolated from the scales of the squirrelfish (*Holocentrus ascensionis*) and cultured on glass coverslips retain their capacity to translocate pigment granules⁹. Mass aggregative movement of the pigment towards the cell centre is rapid and of uniform velocity, while dispersive radial movements are

slow and faltering. When the cell is in the dispersed state, the granules shuttle to and fro in short radial jumps or 'saltations'^{9,10}.

Pressure microinjection¹¹ of erythrophores with $\sim 7 \times 10^{-14}$ l ($\sim 7\%$ of the cell volume) of a 10% sucrose solution at pH 7, containing fluorescein isothiocyanate-labelled bovine serum albumin ($5\text{--}10\text{ mg ml}^{-1}$) results in an initial aggregation of the pigment followed, within 1 min, by a return of normal granule movements. In contrast, the saltatory intracellular motility of erythrophores is rapidly inhibited when the cells are injected with vanadate ($>50\text{ }\mu\text{M}$) in the V(v) oxidation state (Table 1). Microinjection of vanadate at a concentration of $250\text{ }\mu\text{M}$ causes the pigment granules to freeze transiently and then disperse slightly, perhaps because of vanadate enhancement of adenylate cyclase activity¹²; chromatophores characteristically respond to increased intracellular cyclic AMP by dispersing their pigment^{13,14}. Once this initial dispersion has occurred, the shuttling of granules ceases, and the pigment will no longer aggregate, as it usually would, in response to touch by the microinjection needle. Adjacent cells which pulsed in synchrony with the experimental cell before injection will, however, respond by aggregating their pigment when the injected cell is touched (Fig. 1). This result suggests that the injected cell is receiving the stimulus to aggregate and is communicating it to nearby cells, but is incapable of translocating its own pigment in response to the stimulus. Erythrophores injected with 50 or $100\text{ }\mu\text{M}$ V(v) exhibit all the above characteristics except they can occasionally aggregate their pigment in response to touch by the micropipette. However, this aggregation is much slower than that in coupled or control cells and is generally less complete.

When the concentration of vanadate in the injection buffer is increased to $500\text{ }\mu\text{M}$, a cell immediately adjacent to the injected cell often loses its ability to translocate pigment granules, but touching either of these incapacitated cells results in aggregation of other cells previously observed to be coupled to the injected cell. It seems likely that the excess vanadate is passing through the same junctions that serve to couple the cells electrically; the H_2VO_4^- form of V(v) which exists in aqueous solution at neutral pH^{12,15} should pass readily through gap junctions because it is both small and negatively charged. The fact that an uninjected cell is inhibited in its motility illustrates that the injection procedure itself is not responsible for the effects on granule movement noted here. Moreover, as depolarization of the cell membrane is probably the primary signal to aggregate the pigment¹⁶, the immobilized, injected cell must remain alive in order to communicate the aggregation stimulus to adjacent cells.

When the concentration of injected vanadate is decreased to $10\text{ }\mu\text{M}$, an intermediate inhibitory effect is observed. The pigment translocation ceases in 32% of the cells injected while the remainder are affected either transiently or not at all. If the concentration of injected material is decreased to $1\text{ }\mu\text{M}$, all injected cells remain competent to move their pigment.

Kinetic analyses of the effect of vanadate on dynein *in vitro* have demonstrated that its inhibition of the ATPase activity is noncompetitive^{5,6}. Vanadate inhibition of myosin¹⁷ and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ ^{18,19}, on the other hand, is competitive with

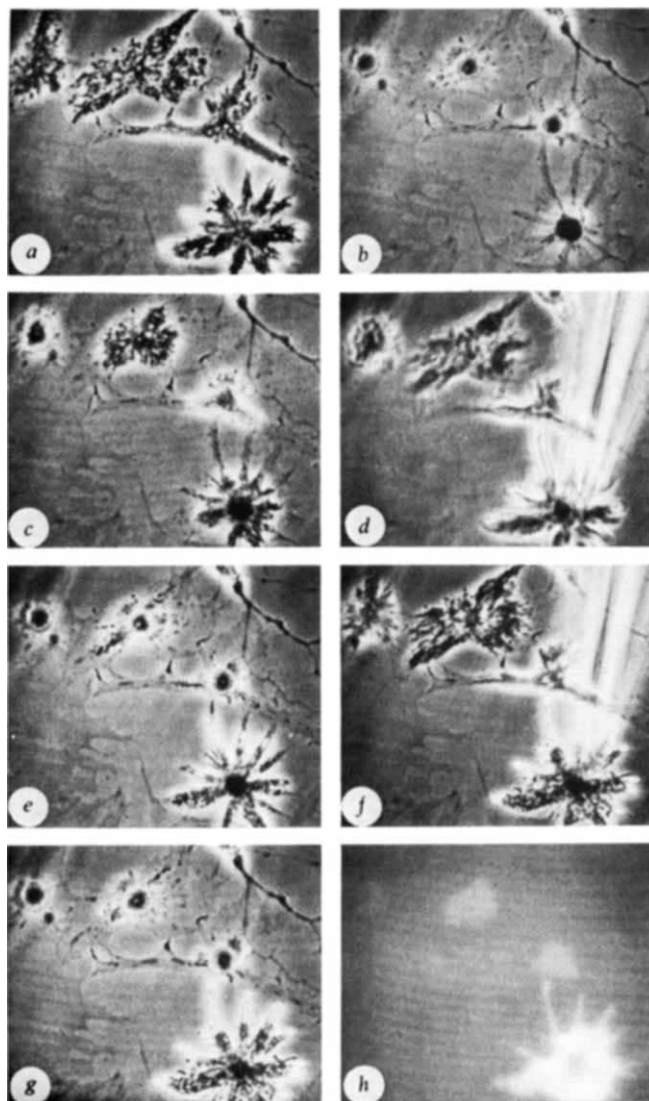


Fig. 1 Phase-contrast micrographs of four erythrophores in culture. The cells are coupled, that is, they disperse (a), aggregate (b) and redisperse (c) their pigment granules in synchrony. In d, the cells are in the dispersed state and one cell is microinjected with vanadate ($250\text{ }\mu\text{M}$) in injection buffer containing fluorescein-labelled bovine serum albumin. The injected cell is rapidly inhibited in its intracellular transport function whereas the three other cells continue to aggregate and disperse their pigment (e, f). Touching the microinjected cell with the micropipette (g) results in aggregation of the three coupled, uninjected cells, but the vanadate-injected cell does not respond to the stimulus with pigment movement (g). A fluorescence image (h) of the four cells illustrates that the cell on the lower right has been successfully injected. Slight autofluorescence of the pigment granules can be seen in the other erythrophores.

Table 1 The effect of vanadate on pigment granule migration

Concentration of vanadate in micropipette (μM)	Approximate intracellular concentration* (μM)	No. of cells injected	Inhibition of transport (no. of cells; % inhibited)
500	37	55	55; 100
250	18.5	29	28; 97
100	7.3	21	21; 100
50	3.7	31	31; 100
10	0.73	22	7; 32
1	0.073	42	0; 0

* Microinjection needles were calibrated by measuring the volume of oil droplets injected into erythrophores.

substrate binding. Injection of a large excess of MgATP (up to 2.5 mM) together with $250\text{ }\mu\text{M}$ vanadate fails to prevent or slow the vanadate effects on pigment granule migration *in vivo*, suggesting that vanadate inhibition of granule motility is probably not due to an effect on any of the numerous ATPases⁶ known to be competitively inhibited by vanadate.

Incubation of erythrophores with another dynein inhibitor, EHNA, at a concentration of $4 \times 10^{-3}\text{ M}$, in culture medium for 20 min also results in cessation of saltatory shuttling of pigment granules. At 10^{-3} M EHNA, a 60–90-min exposure is required before motility is inhibited. The dispersed cells appear to be frozen with regard to particle translocation. Although there is some brownian movement of the pigment granules, they do not

shuttle and no spontaneous aggregation-dispersion cycles occur. Exogenous application of adrenaline stimulates the cells to aggregate their pigment granules, although the response is slowed relative to untreated cells.

EHNA has also been reported to inhibit protein carboxyl-methylase activity by increasing the intracellular concentration of *S*-adenosyl homocysteine²⁰. In view of the fact that calcium appears to be the intracellular signal for aggregation of pigment granules²¹ and calmodulin can be methylated²², we examined the possibility that EHNA inhibition of pigment saltation was due to the inhibition of carboxyl-methylase rather than dynein ATPase. Microinjection of *S*-adenosyl homocysteine at concentrations up to 100 μ M yielded no inhibition of pigment granule shuttling or aggregation and dispersion even after 60 min. As EHNA-induced inhibition of monocyte chemotaxis occurs rapidly when intracellular *S*-adenosyl homocysteine reaches $\sim 6 \times 10^{-18}$ g per cell²⁰, the high concentration injected into erythrocytes ($\sim 3 \times 10^{-15}$ g per cell) should have readily inhibited pigment motion if it relied on carboxyl-methylase activity.

It seems unlikely that either vanadate or EHNA is simply interfering with the general ATP metabolism of the cells

because co-injection of vanadate and ATP affects cells in the same manner as vanadate alone, and the cells treated with either vanadate or EHNA do not ultimately aggregate their pigment as do cells treated with various inhibitors of oxidative phosphorylation¹⁰.

Dynein-like ATPase activity has been isolated from numerous microtubule-based motility systems, from sources as diverse as the sea urchin mitotic apparatus²³, *Saccinobaculus* axostyle²⁴ and human sperm²⁵. Probes for dynein activity have been shown to inhibit anaphase chromosome movement^{7,26}, flagellar beat^{4,8}, microtubule-associated particle transport in fibroblasts²⁷ and, now, erythrocyte pigment movement. Taken together these observations suggest that a dynein-like protein is a force-transducing molecule in microtubule-dependent intracellular motility in general.

This work was supported by a grant from the Muscular Dystrophy Association of America to K.R.P. and a Danforth Fellowship to M.C.B. M.C.B. thanks M. Zavortink for encouragement and instruction in microinjection and M. McNiven for care of the squifreelish. We thank J. R. McIntosh and his team for many helpful discussions and generosity with equipment.

Received 1 November 1981; accepted 15 January 1982.

- Hyams, J. S. & Stebbings, H. in *Microtubules* (ed. Hyams, J. S.) (Academic, New York, 1979).
- Schliwa, M. in *International Cell Biology, 1980-1981* (ed. Schweiger, H. G.) 275-285 (Springer, Berlin, 1981).
- Satir, P. *et al. Cell Motility* **1**, 303-327 (1981).
- Gibbons, I. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2220-2224 (1978).
- Kobayashi, T. *et al. Biochem. biophys. Res. Commun.* **81**, 1313-1318 (1978).
- Shimizu, T. *Biochemistry* **20**, 4347-4354 (1981).
- Cande, W. Z. & Wolniak, S. M. *J. Cell Biol.* **79**, 573-580 (1978).
- Bouchard, P., Penningroth, S. M., Cheung, A., Gagnon, C. & Bardin, C. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1033-1036 (1981).
- Byers, H. R. & Porter, K. R. *J. Cell Biol.* **75**, 541-558 (1977).
- Luby, K. J. & Porter, K. R. *Cell* **21**, 13-23 (1980).
- Graessmann, M. & Graessmann, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 366-370 (1976).
- Simons, T. J. B. *Nature* **281**, 337-338 (1979).

- Geschwind, I. I. *et al. J. Cell Biol.* **74**, 928-939 (1977).
- Beckerle, M. C. & Porter, K. R. *J. Cell Biol.* **91**, 302a (1981).
- Cantley, L. C. *et al. J. biol. Chem.* **252**, 7421-7423 (1977).
- Parker, G. H. *Animal Colour Changes and their Neurohumours*, 122-134 (Cambridge University Press, 1948).
- Goodno, C. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2620-2624 (1979).
- Hansen, O. *Biochim. biophys. Acta* **568**, 265-269 (1979).
- Jorgensen, P. L. & Anner, B. M. *Biochim. biophys. Acta* **555**, 485-492 (1979).
- Pike, M. C., Kredich, N. M. & Snyderman, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3928-3932 (1978).
- Luby, K. J. *Eur. J. Cell Biol.* **22**, 352a (1980).
- Gagnon, C. *Nature* **291**, 515-516 (1981).
- Pratt, M. M., Otter, T. & Salmon, E. D. *J. Cell Biol.* **86**, 738-745 (1980).
- Mooseker, M. S. & Tilney, L. G. *J. Cell Biol.* **56**, 13-26 (1973).
- Baccetti, A. *et al. J. Cell Biol.* **88**, 102-107 (1981).
- Cande, W. Z. *J. Cell Biol.* **91**, 320a (1981).
- Forman, D. S. *J. Cell Biol.* **91**, 414a (1981).

Cyclic AMP induces maturation of trout sperm axoneme to initiate motility

Masaaki Morisawa

Laboratory of Physiology, Ocean Research Institute, University of Tokyo, Nakano, Tokyo 164, Japan

Makoto Okuno

Department of Biology, University of Tokyo, Komaba, Meguro, Tokyo 153, Japan

Cyclic AMP has long been implicated as an activator of sperm motility¹⁻⁵. From more recent experiments using demembrated mammalian and sea urchin spermatozoa^{6,7}, it was concluded that cyclic AMP only increases the motility of the axoneme after it has been initiated by MgATP²⁻. We have now carried out similar experiments using spermatozoa collected from the rainbow trout and demembrated by treatment with the detergent Triton X-100. Our results suggest that in this species, cyclic AMP is required before MgATP²⁻ to trigger maturation of the nonmotile axoneme. Subsequent addition of an energy source then induces motility.

Semen was collected from mature rainbow trout, *Salmo gairdneri*, by inserting a pipette into the sperm duct. The plasma membrane and mitochondria of the sperm were then removed (confirmed in electron micrographs) by mixing at room temperature 1 volume of semen with 20 volumes of the extracting medium containing 0.15 M KCl, 0.5 mM MgSO₄, 5 mM

CaCl₂, 0.5 mM EDTA, 1 mM dithiothreitol (DTT), 2 mM Tris buffer, pH 8.2, and 0.04% Triton X-100 for 30 s in the presence or absence of 0.5 mM cyclic AMP. Removal of the plasma membrane and mitochondria, elimination of the energy supply, and the dispersal of the components essential for axonemal movement in the extracting medium resulted in immotility of the axoneme. Reactivation of axonemal movement was then attempted using the modified method of Gibbons and Gibbons⁸: the demembrated sperm were resuspended in reactivating medium containing 0.15 M KCl, 2 mM MgSO₄, 0.5 mM CaCl₂, 2 mM EGTA, 1 mM DTT, 20 mM Tris buffer, pH 8.2, and 1 mM ATP again in the presence or absence of 0.5 mM cyclic AMP, and were observed by dark-field microscopy with stroboscopic mercury illumination at a magnification of $\times 200$.

During demembration rainbow trout spermatozoa remain quiescent due to the presence of K⁺ in the seminal plasma⁹ or KCl in the extracting medium. When cyclic AMP was absent from both the demembrating and reactivating media, no axonemal movement was observed (Fig. 1a). However, when the demembrated sperm were suspended in reactivating medium containing MgATP²⁻ and cyclic AMP, axonemal movement occurred (Fig. 1b), causing the reactivated sperm to swim at a speed of $103 \pm 12 \mu\text{m s}^{-1}$ ($n = 8$). In contrast, motility of demembrated mammalian and sea urchin sperm is restored in the presence of MgATP²⁻, with no requirement for cyclic AMP⁶⁻⁸. These results suggest that whereas the axoneme of mammalian and sea urchin sperm is already functionally mature and requires only an energy supply to initiate motility, the axoneme of rainbow trout sperm is functionally immature and is converted to the mature state by exposure to cyclic AMP.

When trout spermatozoa were demembrated in the extracting medium (always ATP free) containing cyclic AMP and then

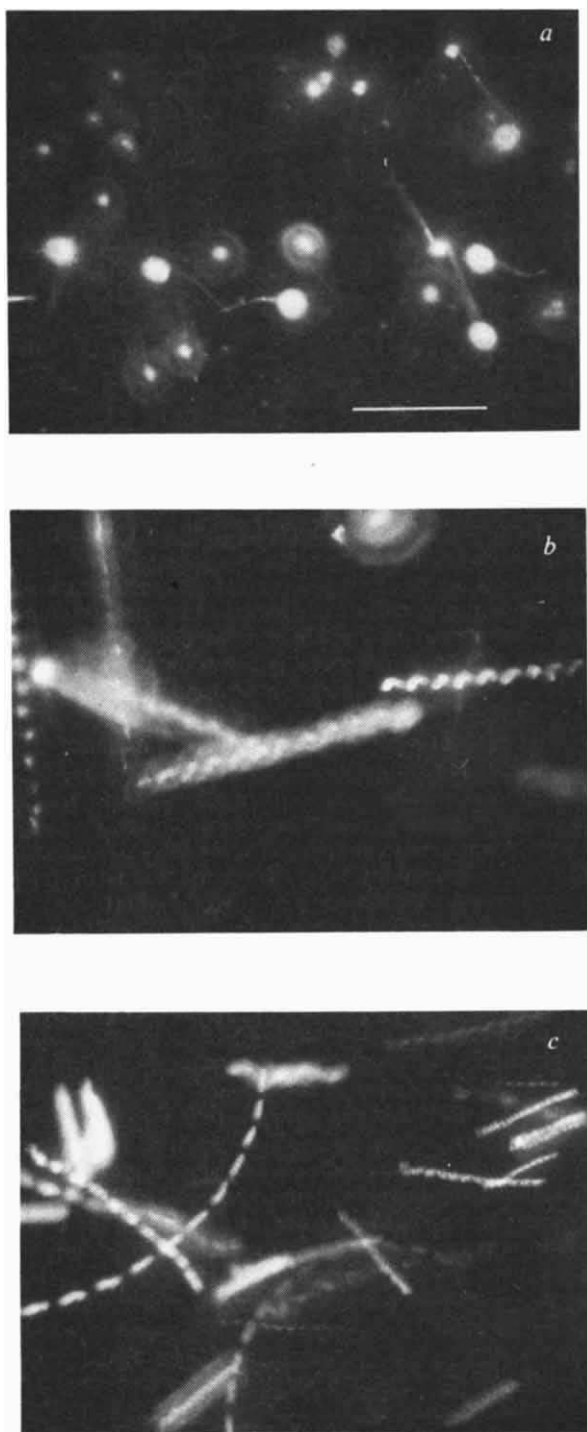


Fig. 1 Dark-field stroboscopic micrographs of tracks of the heads of spermatozoa demembranated by treatment with Triton X-100. In *a*, cyclic AMP was absent from both the extracting and reactivating media; semen was directly immersed in the extracting medium and demembranated; no evidence of movement. In *b*, cyclic AMP was present in the reactivating medium; demembranation of sperm was performed by the same method as in *a*; movement is evident. In *c*, cyclic AMP was absent from both the extracting and reactivating media; semen was diluted in 100 mM NaCl for 30 s and then spermatozoa were demembranated; movement is evident. In *a*, demembranated sperm are immotile; rounded heads and slender axoneme projected from heads are observed. In *b*, *c*, demembranated sperm are motile; the image of tracks of heads which are driven by axonemal movement are observed as a dotted line under stroboscopic illumination; all preparations exhibit straight or curved lines, no rounded heads are present, indicating 100% motility of axonemes. Bar in *a* indicates 50 μm .

placed in reactivating medium containing MgATP^{2-} but no additional cyclic AMP, axonemal movement occurred to give a speed of $78 \pm 9 \mu\text{m s}^{-1}$ ($n = 4$). This can be explained if cyclic AMP becomes bound to the axoneme during the demembranated procedure, and induces its maturation. Axonemal movement can then occur with the addition of the energy source, MgATP^{2-} alone.

In the above experiments Ca^{2+} concentration was $\sim 5 \text{ mM}$ in the extracting medium and 10^{-9} M in the reactivating medium (buffered with EGTA). Even when CaCl_2 was eliminated from both media, axonemal movement occurred, suggesting that Ca^{2+} did not affect the maturation in motility of the axoneme.

During natural spawning, seminal K^+ , which inhibits motility of salmonid sperm in the sperm duct, is diluted by fresh water, allowing sperm motility to occur¹⁰. In the present experiments, the mature axoneme deprived of plasma membrane and mitochondria always displayed motility in reactivating medium containing 0.15 M KCl, suggesting that K^+ ions inhibit the motility of intact (functionally mature) sperm through metabolic or membrane effects, and do not directly inhibit axonemal maturation. When an appropriate volume of semen was diluted to 100 vol in K^+ -free 100 mM NaCl for 30 s (the duration of intact sperm motility of rainbow trout) and then a volume of the sperm suspension was extracted in the absence of cyclic AMP, axonemal movement occurred to give a speed of $100 \pm 32 \mu\text{m s}^{-1}$ ($n = 11$) in reactivating medium containing MgATP^{2-} , again in the absence of cyclic AMP (Fig. 1c). This result offers an explanation for normal initiation of trout sperm motility: the reduction of external K^+ at spawning may affect metabolic or membrane-mediated processes in the intact sperm and thus induce an increase in intraflagellar cyclic AMP concentration. The endogenous cyclic AMP then induces axonemal maturation (as described above; Fig. 1b), and the matured axoneme conveys motility to the sperm, using energy provided by the mitochondria.

Components of the cyclic AMP system, including cyclic AMP, adenylate cyclase, phosphodiesterase, cyclic AMP-dependent protein kinase which phosphorylates the motile apparatus¹¹, and phosphoprotein phosphatase, occur in motile cells, including spermatozoa¹², cilia¹¹ and muscle¹³, of many animals. The ability to isolate an immature motile apparatus from trout sperm provides a useful model for investigating cell motility, especially the molecular site of cyclic AMP action and phosphorylation in sperm motility and the chemistry of interaction between the cyclic AMP system and the motile apparatus in initiation of cell motility generally.

Independently of our present study, activation of demembranated spermatozoa of a tunicate *Ciona* by cyclic AMP has been described¹⁴.

We thank H. A. Bern and C. J. Brokaw for critical evaluation of the manuscript, H. Mohri for encouragement, and the director and staff of Kumagaya Branch of Saitama Fisheries Experiment Station for providing the materials. This work was supported in part by a grant-in-aid from the Japanese Ministry of Education and Itô Foundation for the Advancement of Ichthyology to M. M.

Received 27 October; accepted 21 December 1981.

- Garbers, D. L., Lust, W. D., First, N. L. & Lardy, H. A. *Biochemistry* **10**, 1825–1831 (1971).
- Morton, B., Harrington-Lum, J., Albagli, L. & Jooss, T. *Biochem. biophys. Res. Commun.* **56**, 372–379 (1974).
- Hoskins, D. D., Hall, M. L. & Munsterman, D. *Biol. Reprod.* **13**, 168–176 (1975).
- Wyker, R. & Howard, R. S. *Fert. Steril.* **28**, 108–112 (1977).
- Hoskins, D. D., Brandt, H. & Acott, T. S. *Fedn Proc.* **37**, 2534–2542 (1978).
- Lindemann, C. B. *Cell* **13**, 9–18 (1978).
- Mohri, H. & Yanagimachi, R. *Expl Cell Res.* **127**, 191–196 (1980).
- Gibbons, B. H. & Gibbons, I. R. *J. Cell Biol.* **54**, 75–97 (1972).
- Schlenk, W. Jr & Kahmann, H. *Biochem. Z.* **295**, 283–301 (1938).
- Morisawa, M. & Suzuki, K. *Science* **210**, 1145–1147 (1980).
- Mohri, H. *Biochim. biophys. Acta* **456**, 85–127 (1976).
- Garbers, D. L., Watkins, H. E., Tubbs, D. J. & Kopf, G. S. *Adv. Cyclic Nucleotide Res.* **9**, 583–595 (1978).
- Adelstein, R. S. & Eisenberg, E. A. *Rev. Biochem.* **49**, 921–956 (1980).
- Brokaw, C. J. *Cell Motility* (in the press).

Different mRNAs from the transforming region of highly oncogenic and non-oncogenic human adenoviruses

Anders Virtanen & Ulf Pettersson

Department of Microbiology, The Biomedical Centre, University of Uppsala, Box 581, S-751 23 Uppsala, Sweden

Jean Michel Le Moullec, Pierre Tiollais & Michel Perricaudet

Unité de Recombinaison et Expression Génétique, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

Only a fraction of the genome of adenovirus, a DNA tumour virus, is required for transformation *in vitro*¹⁻⁴. The leftmost 11% of the genome, including the two transcription units E1A and E1B, appear to be sufficient for transformation of rat embryo cells in tissue culture for highly, weakly and non-oncogenic adenovirus serotypes^{5,6}. Various spliced mRNAs, originating from the transforming region, have been identified⁷⁻¹¹ and in some cases the corresponding protein products found¹²⁻¹⁴ but it is unknown what renders some serotypes highly oncogenic for baby hamsters compared with others that are non-oncogenic although able to transform cells *in vitro*. We have studied the mRNAs from the transforming region of the highly oncogenic serotype 12 (Ad12) in the hope of finding novel mRNAs which could account for the oncogenic properties of the virus. We have previously failed to find any differences between the spliced mRNAs in the transcription unit of the E1A region of Ad12 and the corresponding mRNAs from the non-oncogenic type 2 (Ad2)¹⁵. Here we have extended our analysis to the E1B region and show that the highly oncogenic Ad12 encodes mRNAs which have different structures from those encoded by Ad2.

We isolated mRNA from Ad12-infected cells and prepared cDNA copies which were inserted into the pBR322 vector. RNA was extracted¹⁶ from infected A549 cells (a cell line derived from a human oat-cell carcinoma) either 8 h after infection in the presence of 25 µg ml⁻¹ cycloheximide or 16 h post-infection in the absence of drug. To study the temporal expression of genes in the transforming region we prepared ³²P-labelled cDNA copies from the two mRNA preparations, and hybridized¹⁷ these to restriction enzyme fragments of Ad12 DNA. The results (Fig. 1) show that the RNA from the cycloheximide-treated cells hybridized only to the *AccI* H fragment which covers the E1A region of Ad12 (ref. 18). In contrast, RNA prepared 16 h post-infection in the absence of cycloheximide hybridized also to the *AccI* F fragment and to two small fragments that are specific for region E1B (Fig. 1).

Double-stranded cDNA copies of late (16 h) poly(A)-containing RNA were prepared as described elsewhere¹⁰ and inserted in the *PstI* cleavage site of the pBR322 plasmid. Clones having sequences from the E1 region were identified by screening 50,000 colonies by hybridization¹⁹ using ³²P-labelled *EcoRI* C fragment as probe. To distinguish between clones from regions E1A and E1B, the *AccI* restriction fragments, 1, 2 and 3 (see Fig. 1) of the *EcoRI* C fragment were used and 38 E1B-specific clones were identified. Characterization of the clones by cleavage with endonucleases *HindIII*, *SalI*, *XbaI* and *PvuII*, which cleave the Ad12 DNA within region E1B (Fig. 2), showed that they correspond to three classes of spliced mRNA.

The structures of these classes were determined by sequencing parts of representative clones and comparing the results with

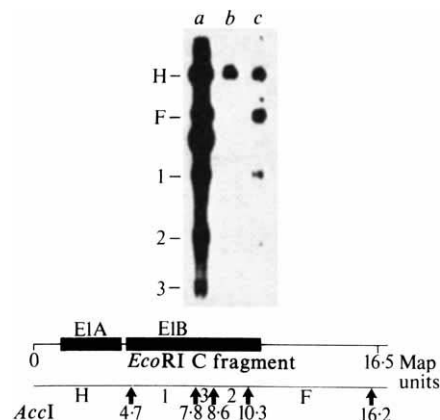


Fig. 1 Kinetic analysis of mRNA expression from the transforming region of Ad12. For this experiment we used a cloned *EcoRI* C fragment of Ad12 DNA comprising the leftmost 16% (M.P., unpublished data). *AccI* cleavage of the cloned fragment results in five fragments of viral DNA which can be used to discriminate between regions E1A and E1B of Ad12. The lower panel shows a restriction enzyme cleavage map for endonuclease *AccI*. Only the leftmost 16.5 map units are included and the fragments are indicated by numbers and letters. The top panel shows hybridization between *AccI* fragments of Ad12 DNA and ³²P-labelled cDNA, prepared using AMV polymerase with oligo(dT) as primer. RNA extracted 8 h after infection (b) and 16 h after infection in the absence of drug (c) was tested. For the analysis a clone containing the *EcoRI* C fragment of Ad12 DNA (0–16.5 map units) was cleaved with *AccI* and the resulting fragments separated on a 1.4% agarose gel before transfer to nitrocellulose¹⁷. ³²P-labelled Ad12 DNA was hybridized to a separate filter strip to visualize all fragments containing viral DNA (a). The band labelled H in the upper panel, contains the *AccI* H fragment of Ad12 plus flanking sequences from the pBR322 plasmid.

the nucleotide sequence determined for the E1B region of Ad12 DNA (ref. 20 and K. Fujinaga *et al.*, personal communication). We thus established the way in which the three mRNA classes are spliced; mRNAs belonging to class I are spliced between nucleotides 3,295 and 3,369 and resemble closely the 22S mRNA from the E1B region of subgroup C adenoviruses²¹. RNA species belonging to class II have two introns; one is identical to the small intron in class I mRNAs, the second is large and located between nucleotides 2,046 and 3,163 in the sequence of Bos *et al.*²⁰. Class III mRNAs also have two introns; one is large and identical to the large intron in class II mRNAs whereas the second is short and located between nucleotides 3,295 and 3,444. Thus the same donor site but different acceptor sites are used for splicing the small intervening sequences present in class II and III mRNAs. All 38 clones in our collection contained the *XbaI* cleavage site at map position 8.9, suggesting that no Ad12 mRNA has the same structure as the 13S mRNA of region E1B in Ad5 (ref. 21), which contains a single large intron.

S₁ nuclease analysis has identified a common poly(A) addition site both for the Ad12 E1B mRNAs¹¹, and the Ad5 E1B

Table 1 Characterization of 38 clones containing sequences from region E1B of Ad12

No. of clones	mRNA species*
4	I
8	II
4	III
22	Unclassified†

* As defined in Fig. 2.

† These clones contained severely truncated cDNA inserts and thus could not be classified unambiguously; most belong to class II and III mRNAs.

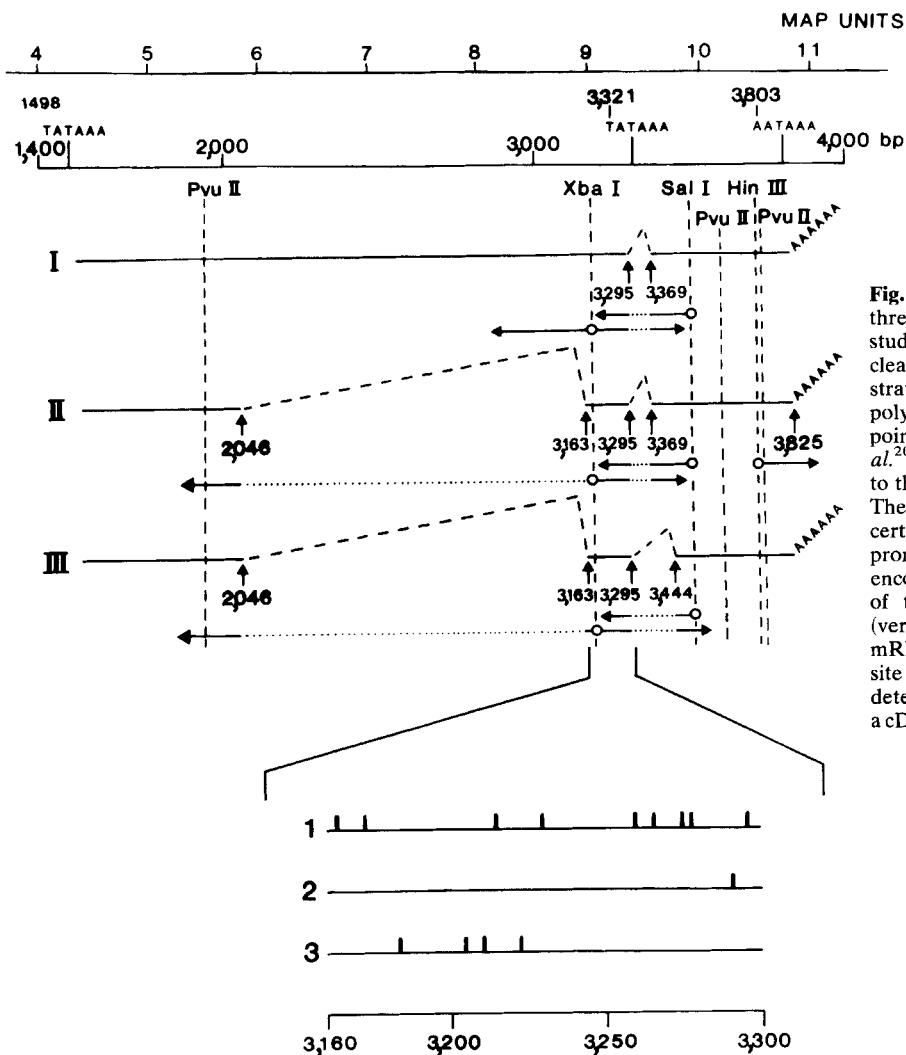


Fig. 2 Schematic drawing of the structure of the three classes of mRNA identified in the present study. The positions of selected restriction enzyme cleavage sites are indicated and the arrows demonstrate how the sequence was determined²⁸ at the poly(A) addition site and at the different splice points. The sequence is indexed according to Bos *et al.*²⁰. The numbers at the exon/intron junctions refer to the first or last nucleotide of the respective exon. The upper part of the figure indicates the positions of certain sequences: TATAAA at position 3,321 is a promoter-like sequence which precedes the sequence encoding the Ad12 polypeptide IX. The lower part of the figure illustrates the termination codons (vertical bars) in the second exon of class II and III mRNAs (data from ref. 20). The poly(A) addition site is located immediately to the left of the sequence determined by Bos *et al.*²⁰ and was determined from a cDNA clone. Nucleotide 3,825 is the first A residue in the poly(A) tail. bp, Base pairs.

mRNAs^{7,8}. By sequencing the ends of a cDNA insert (Fig. 2), the poly(A) addition site in region E1B of Ad12 was located at nucleotide 3,824. It is preceded by the hexanucleotide sequence AAUAAA, which is almost invariably adjacent to poly(A) addition sites in eukaryotic mRNAs²².

The 5' ends of the mRNAs belonging to the three classes cannot be determined from the present study. However, results of S₁ nuclease analysis suggest that all mRNAs from region E1B of Ad12 have a common 5' end located around nucleotide 1,525 in the Ad12 sequence¹¹. There are extensive sequence homologies between this region in Ad12 and the known E1B cap site in the Ad5 genome; furthermore, the sequence TATAAA (the 'Goldberg-Hogness box') lies between nucleotides 1,498 and 1,503.

The coding potential of the three classes of Ad12 mRNA can be predicted if it is assumed that they have common 5' ends around nucleotide 1,525. Bos *et al.*¹² have shown that the E1B region of Ad12 has three major open translational reading frames. One very long reading frame starts with an ATG triplet at position 1,846 and ends with a UGA termination codon at position 3,292 (Fig. 3). This open reading frame has a theoretical coding capacity of molecular weight 54,000 (54 K) and could be used in its entirety only in the long mRNA (class I) with a single splice. When translated from this mRNA the protein would terminate immediately before the splice. The second long open translational reading frame starts with an ATG triplet at position 1,541 and ends at position 2,030 (Fig. 3). From the deduced mRNA structures it is apparent that the latter reading frame can theoretically be used in mRNAs belonging to all three classes (Fig. 3). Class I mRNAs are consequently polycistronic with two

uninterrupted reading frames, possibly encoding two completely different polypeptide chains as has been proposed previously²⁰. These would have molecular weights of 54 K, and 19 K, which roughly correspond to the molecular weights of two

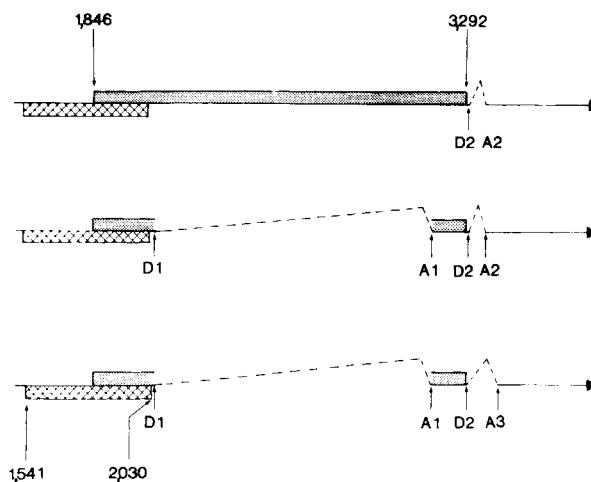


Fig. 3 The two major open translational reading frames in Ad12 as defined by Bos *et al.*²⁰ and their relationship to the structures of the Ad12 E1B mRNAs. The different acceptor (A) and donor (D) sites used to splice E1B mRNAs are indicated.

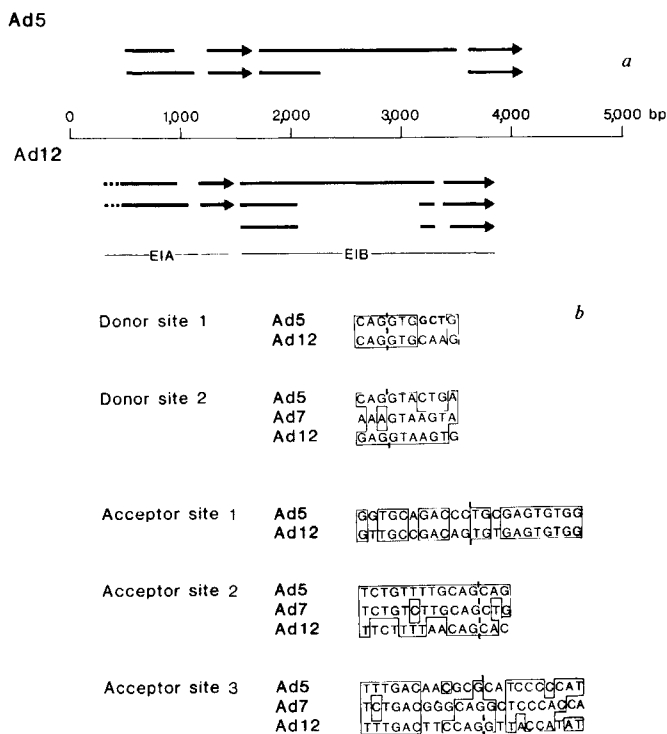


Fig. 4 Comparison of the transforming regions of highly oncogenic and non-oncogenic adenoviruses. *a*, Comparison of the major spliced mRNAs, transcribed from the transforming region of Ad5 and Ad12. *b*, Comparison of nucleotide sequences at different splice junctions in Ad5, Ad7 and Ad12. The donor sites D1 and D2, and the acceptor sites A1, A2 and A3, are defined in Fig. 3. Data for Ad12, Ad5 and Ad7 are from ref. 20, refs 29 and 30, and ref. 26 respectively.

polypeptides identified by translation *in vitro*²³. If translation were initiated at the ATG triplet which precedes the long open reading frame (at position 1,846) in class II and III mRNAs, a polypeptide would be translated from sequences on both sides of the large splice present in these mRNAs (Fig. 3). The product would be a 12 K polypeptide, related to the 54 K polypeptide but lacking 372 internal amino acids. The third open reading frame in region E1B corresponds to polypeptide IX which is translated from an unspliced mRNA²⁰.

The existence of two alternative small introns in class II and III mRNAs is puzzling. Table 1 shows that both types are frequent and hence not the result of occasional incorrect splicing events. As the small exon in class II and III mRNAs contains stop codons in all three possible reading frames (Fig. 2), the coding capacity of class II and III mRNAs is the same.

Figure 4a outlines the mRNAs encoded by the transforming regions of Ad5 and Ad12. The most striking discrepancy between the highly oncogenic Ad12 and the non-oncogenic Ad5 is the existence in the former of two RNA species from region E1B which have a large and a small intron. No equivalent mRNA species have yet been reported from region E1B of subgroup C adenoviruses. The two alternative splices in class II and III mRNAs of Ad12 are both located in the 3'-noncoding part of the mRNA, which is very unusual. Introns in the 3'-noncoding region of cellular mRNAs have been rarely encountered previously²⁴, the only known example besides the E1B mRNAs being the splice present in the mRNA for the small T antigen of simian virus 40 (ref. 25). There are two ways in which the two small splices could influence the properties of class I and II mRNAs: either the structure of the 3'-noncoding region predetermines by some unknown mechanism which of the two possible initiation codons is chosen, thereby generating completely different polypeptides from class II and III mRNAs,

or the different 3'-noncoding regions in class II and III mRNAs alter the stability or rate of transport of the mRNAs. Provided that the difference between oncogenic and non-oncogenic viruses is quantitative rather than qualitative, the unique 3'-noncoding parts of the Ad12 mRNAs could have a decisive role in determining the oncogenic potential of the virus. It should be possible to eliminate specifically the splice which generates the class III mRNAs by site-specific mutagenesis and thus test directly the hypothesis.

The results, summarized in Fig. 4 together with corresponding data from subgroup B and C adenoviruses, make it possible to compare directly nucleotide sequences at the positions where splicing occurs in region E1B of different adenoviruses. Two donor sites, D1 and D2, and three acceptor sites, A1, A2 and A3, for splicing can be identified in the three classes of Ad12 E1B mRNA (see Fig. 3). Figure 4b compares the sequences at the donor and acceptor sites in region E1B of Ad12 with the corresponding sequences in Ad5 and Ad7. The results show that the donor site D1 is well conserved and apparently used to generate a large splice in both Ad5 and Ad12. In addition the second donor site, D2, is well conserved and apparently used to generate a small splice in Ad5, Ad7 (ref. 26) and Ad12. This is in contrast to the leftmost acceptor site A1 which in Ad12 generates the middle splice, not found in Ad5 mRNAs. Sequence homology is observed around the splice site but the A-G dinucleotide is altered in Ad5, which presumably makes it non-functional²⁷. The acceptor site A2 is, as expected, conserved in all three serotypes in contrast to the third acceptor site A3 which is unique for serotypes 7 and 12. Although there is no direct evidence, we can predict from sequence similarities that the E1B mRNAs of Ad7 resemble those of Ad12. The similarity between Ad7 and Ad12 with regard to acceptor and donor sites in region E1B supports the idea that, as was recently proposed¹⁵, the highly and weakly oncogenic adenoviruses have evolved together.

We thank Elsy Johnsen for technical assistance and Marianne Gustafson for secretarial help. The gift of reverse transcriptase from Dr J. Beard is also acknowledged. This investigation was supported by grants from the Swedish MRC, the Swedish Cancer Society, CNRS, INSERM, and Fondation pour la Recherche Medicale.

Received 13 October; accepted 18 December 1981.

- Sharp, P. A., Petterson, U. & Sambrook, J. *J. molec. Biol.* **86**, 709-726 (1974).
- Gallimore, P., Sharp, P. A. & Sambrook, J. *J. molec. Biol.* **89**, 49-72 (1974).
- Graham, F. L. *et al. Cold Spring Harb. Symp. quant. Biol.* **39**, 637-650 (1974).
- Van der Eb, A. J. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 383-399 (1979).
- Van der Eb, A. J., Mulder, C., Graham, F. L. & Houweling, A. *Gene* **2**, 115-132 (1977).
- Shiroki, K., Shimojo, H., Sawada, Y., Uemizu, Y. & Fujinaga, K. *Virology* **95**, 127-136 (1979).
- Berk, A. J. & Sharp, P. A. *Cell* **14**, 695-711 (1978).
- Chow, L. T., Broker, T. R. & Lewis, J. B. *J. molec. Biol.* **134**, 265-303 (1979).
- Kitchingman, G. R. & Westphal, H. *J. molec. Biol.* **137**, 23-48 (1980).
- Perricaudet, M., Akusjärvi, G., Virtanen, A. & Pettersson, U. *Nature* **281**, 694-696 (1979).
- Sawada, Y. & Fujinaga, K. *J. Virol.* **36**, 639-651 (1980).
- Halbert, D. W., Spector, D. J. & Raskas, H. J. *J. Virol.* **31**, 621-629 (1979).
- Esche, H., Mathews, M. B. & Lewis, J. B. *J. molec. Biol.* **142**, 399-417 (1980).
- Johansson, K. *et al. J. Virol.* **27**, 628-639 (1978).
- Perricaudet, M., Le Moulec, J. -M., Tiollais, P. & Pettersson, U. *Nature* **288**, 174-176 (1980).
- Brawermann, G., Mendecki, I. & Lee, S. Y. *Biochemistry* **11**, 637-641 (1976).
- Southern, E. *J. molec. Biol.* **98**, 503-518 (1975).
- Sugisaki, H. *et al. Cell* **20**, 777-786 (1980).
- Grunstein, N. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3964 (1975).
- Bos, J. L. *et al. Cell* **27**, 121-131 (1981).
- Perricaudet, M., Le Moulec, J. -M. & Pettersson, U. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3778-3782 (1980).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
- Jochimsen, H., Daniels, G. S. G., Lupker, J. H. & van der Eb, A. J. *Virology* **105**, 551-563 (1980).
- Breathnach, R. & Chambon, P. A. *Rev. Biochem.* **50**, 349-383 (1981).
- Reddy, V. B., Gosh, P. K., Lebowitz, P., Piatak, M. & Weissman, S. M. *J. Virol.* **30**, 279-296 (1979).
- Dijkema, R., Maat, J., Dekker, B. M. M., van Ormondt, H. & Boyer, H. W. *Gene* **13**, 375-385 (1981).
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, G. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853-4857 (1978).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Maat, J., Van Beveren, C. P. & Van Ormondt, H. *Gene* **10**, 27-38 (1980).
- Maat, J. & van Ormondt, H. *Gene* **6**, 75-90 (1980).

Site-specific mutagenesis by error-directed DNA synthesis

Richard A. Zakour & Lawrence A. Loeb*

The Joseph Gottstein Memorial Cancer Research Laboratory, University of Washington, Department of Pathology SM-30, Seattle, Washington 98195, USA

An important experimental strategy for defining functional regions of a DNA sequence involves deleting or mutating particular sequences *in vitro* and observing the effects of the changes in a functional assay. We have approached the problem of making specific changes at defined sites in a nucleotide sequence by attempting to make use of the mistakes made by DNA polymerase enzymes while copying a sequence. Our approach is to initiate DNA synthesis from an isolated restriction fragment, and to elongate the fragment to a designated position by sequentially adding complementary nucleotides, then to incorporate the non-complementary nucleotide using an error-prone DNA polymerase followed by continuation of faithful synthesis. Here we demonstrate by DNA sequence analysis that both transitions and transversions can be produced at a designated position on the DNA template using this method.

DNA polymerases from prokaryotes contain a proofreading 3' → 5' exonuclease activity that enhances their accuracy in catalysis¹. This associated exonuclease excises any non-complementary nucleotides at a primer terminus before continuing polymerization. In contrast, DNA polymerases from RNA tumour viruses and eukaryotes incorporate non-complementary nucleotides at higher frequency during copying of synthetic polynucleotides² and natural DNA³. The non-complementary nucleotides incorporated are invariably found to be present as single-base substitutions. Because the frequency of misincorporation by DNA polymerases is proportional to the ratio of non-complementary to complementary nucleotides in the reaction mixture, it should be possible to force a particular non-complementary nucleotide to be incorporated by carrying out the reaction in the presence of only that deoxynucleoside triphosphate. Incorporation of a series of non-complementary nucleotides is unlikely, as it would result in separation of the primer terminus from the template and diminish further synthesis.

To study the feasibility of using nucleotide misincorporation during *in vitro* DNA synthesis for site-specific mutagenesis, we chose to copy single-stranded bacteriophage ΦX174 circular DNA template containing an amber mutation. The frequency of mutagenesis can be determined from the reversion frequency of the copied DNA in transfection assays and the specificity of mutagenesis can be documented by sequence analysis^{4,5}.

Restriction endonuclease cleavage fragment, *HpaI* 2 (ΦX174 positions 31–1,294)⁶ from ΦX174 RF DNA, was hybridized to ΦX174 amber (*am*)18 template DNA. The 3' OH terminus of this restriction fragment is located seven nucleotides from the change that causes the amber 18 mutation (*am* 18 contains a T in place of C at ΦX174 position 23)⁷. We first elongated the 3' OH terminus of the restriction fragment primer using *Escherichia coli* DNA polymerase I (Pol I) with only dATP and dCTP. These nucleotides are complementary to the first three template positions beyond the 3' OH primer terminus (Fig. 1). After isolating template containing the elongated primer, a second series of reactions was carried out using avian myeloblastosis virus (AMV) DNA polymerase. Three different combinations of deoxynucleoside triphosphates were used (Fig. 1). After incubation of the reaction mixtures for 5 min at 37 °C in the

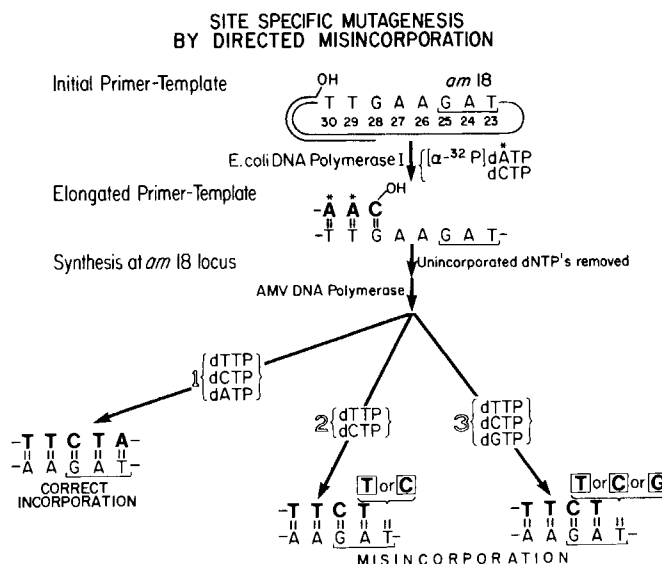


Fig. 1 Steps involved in primer elongation and site-specific mutagenesis. Primer elongation with complementary nucleotides: *HpaI* 2 restriction endonuclease fragment was hybridized to single-stranded ΦX174 *am*18 DNA. Limited elongation of the 3' OH terminus of the primer was carried out with Pol I. The reaction contained 3 μg of primed template, 5 μg of Pol I, 1 μM [α -³²P]dATP, 1 μM dCTP, 5 mM MgCl₂, 2 mM dithiothreitol and 20 mM Tris-HCl (pH 8.0) in a volume of 100 μl. Incubation was at 37 °C for 5 min. The reaction was stopped by adding EDTA to a final concentration of 10 mM; the mixtures were heated to 65 °C for 10 min and extracted once with an equal volume of redistilled phenol equilibrated with 20 mM Tris-HCl (pH 8.1) containing 1 mM EDTA and 100 mM NaCl. To separate the DNA from the unincorporated nucleotides, the aqueous phase was loaded onto a Sephadex G-100 column (1 × 60 cm) and the column eluted with 0.15 M KCl, 50 mM Tris-HCl (pH 7.8). Fractions (1 ml) were collected and aliquots (20 μl) processed for determination of acid-insoluble radioactivity. The peak fractions were pooled, 1/10th volume of 3 M sodium acetate was added, and the DNA was precipitated with EtOH. The DNA was pelleted, washed and resuspended in 100 μl of 10 mM Tris-HCl (pH 8.0) containing 0.1 mM EDTA. Synthesis at *am*18 locus: A second series of *in vitro* DNA reaction mixtures each contained 0.1 μg *am*18 ΦX174 DNA primed with the elongated *HpaI* 2 restriction fragment, 1 U of AMV DNA polymerase, 50 mM Tris-HCl (pH 8.0), 5 mM MgCl₂ and 2 mM dithiothreitol in a volume of 50 μl; they were incubated at 37 °C for 5 min. All reaction mixtures contained 10 μM dTTP and dCTP to elongate the primer correctly to position 23 of the *am*18 locus. Various combinations of the other deoxynucleoside triphosphates, each at a concentration of 10 μM, were included in individual reactions as indicated. (1) Correct incorporation is achieved by including the three nucleotides complementary to the *am*18 locus, dTTP, dCTP and dATP. (2) The presence of only dTTP and dCTP during directed misincorporation synthesis will allow the misinsertion of either T or C at position 23 of the *am*18 locus. (3) By including dTTP, dCTP and dGTP during directed misincorporation synthesis, it should be possible to obtain T, C, or G misinsertion at position 23. In experiments performed in this manner, G substitution at position 23 was the only change observed among seven revertants (Table 2). Synthesis with AMV DNA polymerase was followed by *in vitro* DNA synthesis with all four deoxynucleoside triphosphates present to seal the misincorporated nucleotide into the newly replicated DNA. Reactions were stopped by adding 0.1 vol of 100 μM EDTA.

presence of the specified nucleotides, all four deoxynucleoside triphosphates were added. Synthesis was allowed to proceed for an additional 5 min, so that any misincorporated nucleotides on termini would be protected from excision by the subsequent incorporation of the next complementary nucleotide. EDTA was added to stop synthesis, and the copied DNA was transfected into *E. coli* spheroplasts. Synthesis was completed by the *E. coli* replicating complex *in vivo*. We determined the reversion frequency of the *am*18 mutation by measuring the number of infective centres^{4,8,9}.

When dATP, the nucleotide complementary to position 23 in the *am*18 region, was present in addition to dCTP and dTTP during DNA synthesis using AMV DNA polymerase, the reversion frequency of the newly synthesized DNA product in the transfection assay was fourfold higher than that of the

* To whom correspondence should be addressed.

uncopied primed template (Table 1, lines 1, 2). This increase, which is a function of the error-prone nature of AMV DNA polymerase², provides the background reversion frequency (4×10^{-4}). In both the nucleotide substrate combinations studied (Table 1, lines 3, 4), in which dATP was excluded from the reaction, the reversion frequencies were increased 100-fold. These higher reversion frequencies were presumably due to the substitution by AMV DNA polymerase during the *in vitro* synthesis of one of the non-complementary nucleotides for the missing complementary dATP.

To determine more precisely the efficiency of non-complementary nucleotide misinsertion, we measured the reversion frequency of a population that included only copied DNA molecules. Following mutagenesis with AMV DNA polymerase, synthesis was allowed to continue completely around the Φ X174 template using all four deoxynucleoside triphosphates, and a double-stranded closed circular molecule was formed using DNA ligase; nuclease S₁ was then used to hydrolyse the single-stranded molecules. In initial experiments, the reversion frequency of molecules copied past the *am*18 locus with dCTP, dGTP and dTTP (Table 1, line 4) was increased to 33%. Thus, enrichment for copied DNA molecules should provide designated mutants, even in the absence of a biological procedure for mutant selection.

To show that the increased reversion frequencies were a result of single-base substitutions, we sequenced the DNA of individual revertants. The nucleotide in position 23 of the minus strand was confirmed to be G in wild-type DNA and A in *am*18 DNA (Table 2, line I). In six spontaneous revertants, G substituted for A at position 23. Among the eight revertant DNAs, from reactions in which only dCTP and dTTP were present, seven contained cytosine and one contained thymidine at position 23 (Table 2, line II-1). These changes were probably a result of T-C and T-T mispairings during *in vitro* DNA synthesis. When dGTP was included in the synthesis reaction with dCTP and dTTP, all seven of the revertant DNAs contained G at position 23 (Table 2, line II-2), indicating a T-G mispairing. Thus, all three misincorporations occurred opposite this template thymidine during the *in vitro* DNA synthesis. The substitutions at Φ X174 position 23 were the only changes that we observed in the DNA sequence analysis. This observation suggests that the reaction conditions yielded misinsertions predominantly at the predetermined template position, the first template position the polymerase encountered that lacked a complementary nucleotide.

The patterns of nucleotide misinsertion observed in these experiments indicate that AMV DNA polymerase is capable of catalysing pyrimidine-pyrimidine as well as purine-pyrimidine mispairings. Based on the revertant DNAs that we screened, when dATP was the only nucleotide deleted during synthesis, the presence of dGTP resulted in only T→C transitions, produced by T-G mispairing during synthesis. Despite the presence of both dCTP and dTTP, we observed no transversions

Table 2 DNA sequence analysis of revertants obtained from copying *am*18 DNA with AMV DNA polymerase in the presence of defined deoxynucleoside triphosphate substrates

	Sequence at position [minus (copied) strand]			No. of DNAs sequenced
	25	24	23	
I. Controls				
1. Wild type	C	T	G	2
2. <i>am</i> 18	C	T	A	2
3. Spontaneous revertants	C	T	G	6
II. Revertants, added dNTPs				
1. dTTP, dCTP	C	T	C	7
	C	T	T	1
2. dTTP, dCTP, dGTP	C	T	G	7
	C	T	C	0
	C	T	T	0

Viral template DNA which was wild type for *am*18 was isolated from *am*3 Φ X174 bacteriophage (a gift of J. Weiner and A. Kornberg) and DNA from *am*18 was isolated from Φ X174 *am*18 bacteriophage (a gift of P. Weisbeek) as previously described^{4,8}. Spontaneous *am*18 revertants were derived from non-permissive infective centres obtained from transfection of uncopied *am*18 molecules. Revertant DNAs were isolated from plaque-purified non-permissive infective centres obtained from *am*18 DNAs copied with AMV DNA polymerase as described in Table 1 legend. Isolation of revertant phage and preparation of viral template DNA for sequencing were performed as described elsewhere⁵. The DNA sequence from Φ X174 positions 1–26 of wild type, *am*18 and revertants of *am*18 were determined by the chain terminator method of Sanger *et al.*²³. A preparation containing all three fragments from a *Hpa*I restriction digest of Φ X174 RF DNA was used to prime each single-stranded template at a ratio of 1:1. The sequencing reactions contained *E. coli* DNA polymerase I at an enzyme to template ratio of 10:1 and dideoxy- to deoxynucleotides at a ratio of 200:1. After completing the polymerization reactions, restriction endonuclease *Taq*I was added at a ratio of 1 unit per μ g of DNA and the sample was incubated at 65°C for 60 min. The products of the sequencing reaction were displayed on a 12% polyacrylamide gel. The DNA sequence in the *am*18 region (Φ X174 position 23–25) can be unambiguously read because all synthesis initiated at the other *Hpa*I restriction fragment primers are in larger *Taq*I digestion fragments that do not migrate into the readable portion of the gel.

in DNAs derived from these copying conditions, even though we have observed that substitutions of these nucleotides at position 23 are genetically viable. Thus, it seems that T-G mispairing may be more likely to occur than either T-T or T-C mispairing. This observation agrees with investigations by others in suggesting that purine-pyrimidine mispairing should be favoured over pyrimidine-pyrimidine mispairing^{10,11}.

From these experiments, we propose that a general method for site-specific mutagenesis can be developed using directed misincorporation during DNA polymerization. The most direct application is with single-stranded or gapped DNA templates. Site-specific mutagenesis of double-stranded DNA can also be performed by this method when cloned into a bacteriophage M13 single-stranded vector. The first step of the procedure is to elongate a primer terminus to reach a pre-selected site by using a prokaryotic DNA polymerase and different nucleotide combinations. This can be done in a series of reactions to achieve any required degree of elongation¹². Pol I is particularly suitable for the elongation steps because it is highly accurate with natural DNA templates, the error rate being about 1 in 10^6 (ref. 5). For incorporation of a single non-complementary nucleotide, AMV DNA polymerase offers several advantages. First, it lacks a 3'→5' exonuclease activity and is thus unable to excise a non-complementary nucleotide once incorporated¹³. Second, because AMV DNA polymerase can add complementary nucleotides into terminal single-base mismatches, misinsertions are sealed in place by subsequent correct nucleotide incorporations. Most importantly, in the presence of all complementary nucleotides, the frequency of misincorporation by AMV DNA polymerase is much higher than by polymerases containing proofreading activities^{2,9}. Thus, in copying a natural DNA template, AMV DNA polymerase can be forced to be even more error prone by omitting any complementary nucleotide substrate (that is, an 'infinite pool bias') from the *in vitro* synthesis reaction. An alternative approach is to use Pol I in the presence of only a designated non-complementary deoxynucleoside thiotriphosphate. This analogue is incorporated with

Table 1 Reversion frequency of *am*18 DNA copied by AMV DNA polymerase in the presence of various combinations of deoxynucleoside triphosphate substrates

Added dNTPs	Non-permissive infective centres ($\times 10^3$) (revertants)	Total infective centres ($\times 10^5$)	Reversion frequency ($\times 10^{-4}$)
1. None	3.5	3.41	1
2. dTTP, dCTP, dATP	3.5	0.79	4
3. dTTP, dCTP	291	0.76	383
4. dTTP, dCTP, dGTP	182	0.45	404

The *am*18 template DNA with an elongated *Hpa*II primer was copied with AMV DNA polymerase as described in Fig. 1 legend in the presence of the indicated deoxynucleoside triphosphates. The transfection procedure for determining the number of infective centres and the reversion frequency has been described previously^{4,8}.

normal base-pairing properties but not excised by the 3'→5' activity of the polymerase¹⁴.

Site-specific mutagenesis at a predetermined position on a DNA molecule has become more feasible with the accumulation of DNA sequence information from a variety of genes. To date, three methods have been used to introduce specific base mutations into DNA. The first requires the chemical or enzymatic synthesis of an oligonucleotide that contains a designated single-base change¹⁵⁻¹⁷. The oligonucleotide is hybridized onto a template and extended into a biologically active molecule by *in vitro* DNA synthesis. Even though this may be the most exacting method of site-specific mutagenesis, each different substitution requires a unique oligonucleotide of sufficient length to form a stable hybrid with a specific region of the template^{15,16,18}. In the second approach, nucleotides in a DNA template are modified to promote mispairing during *in vitro* gap repair synthesis. For example, sodium bisulphite has been used to deaminate specific template cytosines which thereafter will pair with adenine instead of guanine during DNA synthesis to produce C:G→T:A transitions^{19,20}. With the third method, base substitutions are produced by misincorporation of a nucleotide analogue in place of a complementary nucleotide during *in vitro* nick translation. For example, by substituting N⁴-hydroxy-dCTP for dTTP, A:T→G:C transitions were produced^{21,22}. These latter two approaches are restricted to a single type of base substitution. Moreover, it is difficult to introduce the substitution only at a single designated site.

In addition to incorporating a non-complementary nucleotide at a particular site, the method outlined here can be used to generate a 'library' of mutations at a particular region. A restriction endonuclease fragment can be sequentially elongated to produce a new set of fragments of increasing length that can be isolated from a polyacrylamide gel. Each elongated restriction endonuclease fragment can then be used as primer for incorporating a non-complementary nucleotide. A set of elongated fragments isolated from a single gel could be used to produce a series of substitutions throughout a specific region of a gene.

We thank Drs T. A. Kunkel, R. Monnat and R. M. Schaaper for useful discussions and E. James for technical assistance. The procedure for isolating only copied ΦX174 DNA molecules was provided by T. A. Kunkel before publication. This investigation was supported by NIH grants CA-24845 and CA-24998 and NSF grant PCM 76-80439. R.A.Z. is a postdoctoral fellow of the National Institute of General Medical Sciences (GM-07115).

Received 20 October 1981; accepted 8 January 1982.

1. Brutlag, D. & Kornberg, A. *J. biol. Chem.* **247**, 241-248 (1972).
2. Battula, N. & Loeb, L. A. *J. biol. Chem.* **249**, 4086-4093 (1974).
3. Kunkel, T. A. & Loeb, L. A. *Science* **213**, 765-766 (1981).
4. Kunkel, T. A. & Loeb, L. A. *J. biol. Chem.* **254**, 5718-5725 (1979).
5. Kunkel, T. A. & Loeb, L. A. *J. biol. Chem.* **255**, 9961-9966 (1980).
6. Sanger, F. *et al. J. molec. Biol.* **125**, 225-246 (1978).
7. Smith, M., Brown, N. L., Air, G. M., Barrell, B. G. & Coulson, A. R. *Nature* **265**, 702-705 (1977).
8. Weymouth, L. A. & Loeb, L. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1924-1928 (1978).
9. Gopinathan, K. P., Weymouth, L. A., Kunkel, T. A. & Loeb, L. A. *Nature* **278**, 857-859 (1979).
10. Topal, M. D. & Fresco, J. R. *Nature* **263**, 285-289 (1976).
11. Fersht, A. R. & Knill-Jones, J. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4251-4255 (1981).
12. Donelson, J. E. & Wu, R. J. *J. biol. Chem.* **247**, 4654-4660 (1972).
13. Battula, N. & Loeb, L. A. *J. biol. Chem.* **251**, 982-986 (1976).
14. Kunkel, T. A., Eckstein, F., Mildvan, A. S., Kopitz, R. M. & Loeb, L. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6734-6738 (1981).
15. Hutchison, C. A. *et al. J. biol. Chem.* **253**, 6551-6560 (1978).
16. Razin, A., Hirose, T., Itakura, K. & Riggs, A. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4268-4270 (1978).
17. Bhanot, O. S., Kahn, S. A. & Chambers, R. W. *J. biol. Chem.* **254**, 12684-12693 (1979).
18. Astell, C. R. & Smith, M. *Biochemistry* **11**, 4114-4120 (1972).
19. Shortle, D. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2170-2174 (1978).
20. Shortle, D., Pipas, J., Lazarowitz, S., DiMaio, D. & Nathans, D. in *Genetic Engineering, Principles and Methods* Vol. 1 (eds Setlow, J. K. & Hollaender, A.) 73-92 (Plenum, New York, 1979).
21. Müller, W., Weber, H., Meyer, F. & Weissmann, C. *J. molec. Biol.* **124**, 343-358 (1978).
22. Weissmann, C., Nagata, S., Taniguchi, T., Weber, H. & Meyer, F. in *Genetic Engineering, Principles and Methods* Vol. 1 (eds Setlow, J. K. & Hollaender, A.) 133-150 (Plenum, New York, 1979).
23. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).

No evidence for 'stress' α -globin genes in chicken

Walter Knöchel*, Burghardt Wittig*,
Stephanie Wittig*, Maliyakal E. John*,
Ulrich Grundmann*, Walter Oberthür†,
Jasminka Godovac† & Gerhard Braunitzer†

* Institut für Molekularbiologie und Biochemie,
Freie Universität Berlin, D-1000 Berlin 33, Arnimallee 22, FRG
† Max-Planck-Institut für Biochemie, Abteilung Proteinchemie,
D-8033 Martinsried bei München, FRG

Phenylhydrazine-induced anaemia in goats and certain sheep carrying a β^A -globin gene, leads to a switch from adult β^A -globin mRNA to β^C mRNA synthesis¹. The β^C globin chains are also transiently produced late in fetal life. The nucleotide sequence of selected fragments of cloned goat β^C globin gene² corresponds to the amino acid sequence determined some years ago³. A similar reactivation of globin genes has been reported for the baboon, *Papio cynocephalus*, where anaemic stress also leads to the production of fetal haemoglobin⁴. In the case of anaemic chicken, results of cDNA cloning and nucleotide sequencing experiments suggested that a novel type of α -globin gene, the 'stress' gene, is expressed⁵. The existence of such a gene was mainly supported by the fact that the cDNA nucleotide sequence did not agree with any known chicken α -globin chain amino acid sequence. In contrast, we now report results, including a reinvestigation of the α^A -globin amino acid sequence, which reveal that in the anaemic chicken, the α^A gene is expressed, and there is no evidence for a specific stress α -globin gene.

Reactivation of genes due to physiological stress, such as induction of anaemia, is an important concept in the regulation of gene activity. The difference between the stress gene reported in chicken and the fetal genes of animals mentioned above is that the protein product of the stress gene could not be identified among the adult (α^A , α^D and β) or embryonic (π , π' , α^A , α^D , ϵ and ρ) globin chains⁶. Therefore, this switch seems not to represent a reactivation of a fetal-type gene. For further analysis of globin-coding sequences expressed in anaemic chicken, globin mRNA was isolated from polychromatic erythrocytes, transcribed into cDNA and cloned in bacterial plasmids⁷. Hybrid arrested cell-free translation from nine randomly selected clones demonstrated that six clones contain sequences coding for α chains and three sequences coding for β chains. The integrated sequence of one of the plasmids containing an α -globin sequence (p α^A HG3), which is 581 nucleotides long, was cleaved by *Pst*I, and after digestion with different restriction enzymes (see Fig. 1) the various fragments were submitted to

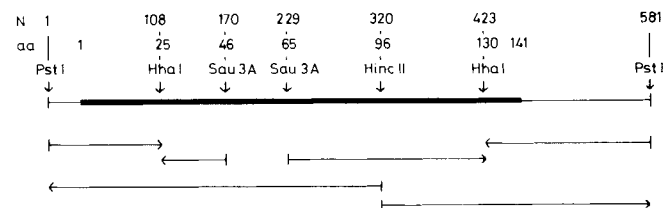


Fig. 1 Restriction map and sequencing strategy of the cloned nucleotide sequence cleaved by *Pst*I from plasmid p α^A HG3. N refers to the number of nucleotides and aa to the number of amino acids, after which cleavage by restriction endonucleases occurs. The ends of the cloned sequence are tailed by ~20 G-C base pairs, originating from the G-C insertion technique in plasmid pBR322⁷. Restriction fragments, as shown by arrows, were subjected to sequence analysis (see Fig. 3).

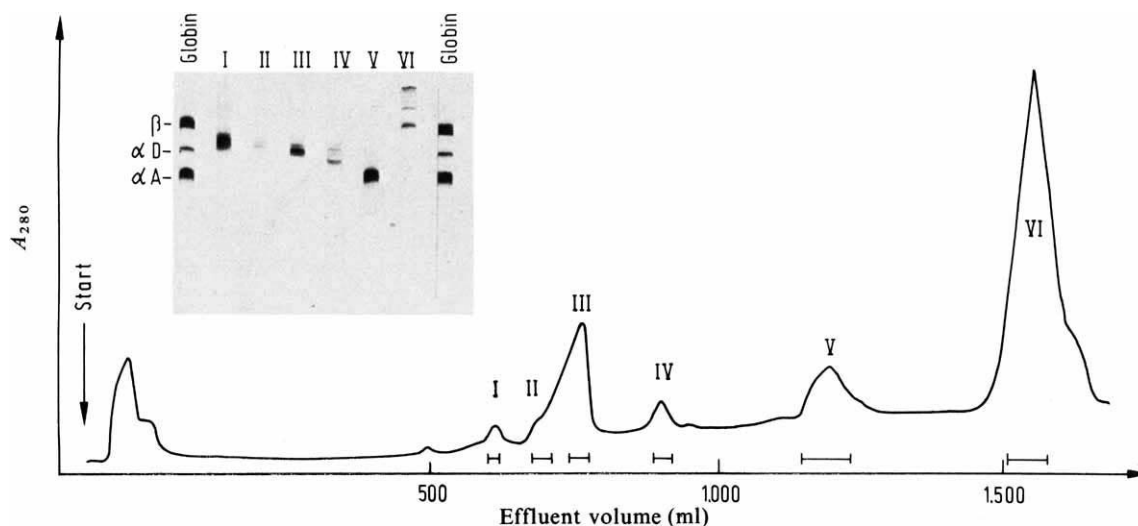


Fig. 2 Carboxymethyl-cellulose chromatography and Triton X-100 electrophoresis of chicken globin chains. Globin chains were isolated from erythrocyte lysates by precipitation of the 100,000g supernatant with acidic acetone and submitted to CM-cellulose (CM 52, Whatman) chromatography¹⁵. The eluting material was further analysed by Triton X-100 polyacrylamide gel electrophoresis¹⁶ (inset).

nucleotide sequence analysis. The resulting nucleotide sequence includes the entire protein-coding region (see Fig. 3) and is very similar to sequences previously reported independently by three other groups⁸⁻¹⁰. The amino acid sequence derived from our cloned structural α -globin gene is identical with that of one of these three groups⁹, although there are variations in the third bases of six codons. Our sequence differs at amino acid positions 110 and 121 or at positions 92, 93, 94 and 120, respectively, from the sequences published by the other two groups^{8,10}. Although the origin of these differences is not known, it is concluded that our cloned sequence, as well as those reported by the above groups⁸⁻¹⁰, represent the processed transcripts of the same gene coding for the major α -globin chain component of polychromatic erythrocytes in anaemic chicken.

A comparison of the amino acid sequence derived from our nucleotide sequence showed differences from embryonic π or π' chains¹¹ at 60 positions, from the adult α^D chain¹² at 61 positions and from the adult α^A chain¹³ at 22 positions. Thus it seemed that the cloned globin-coding sequence from anaemic chicken differs from the structural genes coding for these α -type chains. However, we re-examined the data for the α^A chain because (1) amino acid composition data^{6,14} correspond better to our nucleotide-derived sequence than to the published sequence¹³, and (2) a comparison of this α^A chain with the amino acid sequence of the major α -chain component of phylogenetically related animals, such as pheasant (see Fig. 3), revealed a rather large number of 21 exchanges.

For this reinvestigation, the native globin of adult chicken was separated by CM-cellulose (CM 52) chromatography¹⁵, and the different peaks were further analysed by Triton X-100 polyacrylamide gel electrophoresis¹⁶ (Fig. 2). Peak V was identified as the major α component, α^A . The relatively small absorbance can be explained by the fact that the α^A chains contain no tryptophan in contrast to four in the β chains¹⁷ and one in the α^D chains¹². The CM elution profiles and the electrophoretic patterns of globin chains from mature and immature cells were identical (data not shown). Amino acid analysis was performed on complete chains, tryptic peptides and the C-terminal fragment generated by acidic hydrolysis of the Asp-Pro peptide bond¹⁸ using a Beckman model 121 C amino acid analyser. The chains and peptides were degraded automatically in a Beckman 890 B sequencer or in a Beckman 890 C sequencer with programs as reported elsewhere¹⁵. The phenylthiohydantoin derivatives of the amino acids were identified by TLC and HPLC¹⁹.

There were no differences in the peptide patterns, amino acid analysis of chains and peptides as well in the sequences of few peptides that reportedly differ in α^A and stress gene products,

regardless of whether α chains were isolated from mature or immature blood cells. Figure 3 shows the sequence, which corresponds to that derived from the nucleotides. Moreover, it is obviously homologous to the major α -chain component of the closely related species pheasant (only five differences at positions α 28, α 34, α 38, α 70 and α 77), thus supporting the validity of our results. Our sequence data differ at 22 amino acid residues from that reported more than 10 years ago¹³. We are unable to identify such a chain in normal or anaemic chicken.

These findings demonstrate that the cloned sequence actually represents the structural gene sequence of the chicken α^A -globin chain. There is no reason to postulate a specific type of α -globin gene which is expressed during anaemic stress⁵. There was also no evidence of such a stress gene when the genomic arrangement of the complex α -globin gene family was investigated²⁰. Moreover, a direct comparison of α chains from erythrocytes of uninduced animals and from polychromatic erythrocytes of anaemic animals (7 days after first phenylhydrazine injection) revealed no differences. Whether anaemia for longer periods induces reactivation of embryonic genes similar to that found in sheep and goats¹ cannot be excluded and remains to be elucidated. However, it is clear that the cloned α -globin cDNA

No. aa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
a	AACC	AUG	GUG	GUC	UCC	GCU	GCU	GAC	AAG	AAC	AAC	GUC	AAG	GGC	AUC	UCC	ACC	AAA	AUC	GCC	GCC	CAU	GCU	GAG
b	VAL	LEU	SER	ALA	ALA	ASP	LYS	ASN	ASN	VAL	LYS	GLY	ILE	PHE	THR	LYS	ILE	ALA	GLY	HIS	ALA	GLU	GLU	
c	VAL	LEU	SER	ALA	ALA	ASP	LYS	ASN	ASN	VAL	LYS	GLY	ILE	PHE	THR	LYS	ILE	ALA	GLY	HIS	ALA	GLU	GLU	
d	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
a	UAA	GGC	GCG	GAG	ACC	CUA	GAA	AGG	AUG	UUC	ACC	ACC	UAC	CCC	CCA	ACC	AAG	ACC	UAG	UCC	CCC	CAC	UUC	GAU
b	TYR	GLY	ALA	GLU	THR	LEU	GLU	ARG	MET	PHE	ILE	THR	TYR	PRO	THR	LYS	THR	TYR	PHE	PRO	HIS	PHE	ASP	LEU
c	TYR	GLY	ALA	GLU	THR	LEU	GLU	ARG	MET	PHE	ILE	THR	TYR	PRO	THR	LYS	THR	TYR	PHE	PRO	HIS	PHE	ASP	LEU
e	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
a	UCA	CAC	GCG	UCC	CCU	CAA	AUC	AAG	GGG	CAC	GCG	ANG	AAG	GUA	GCU	GCC	UUG	AAG	GAG	GCU	GCC	AAC	CAC	AAU
b	SER	HIS	GLY	SER	ALA	GLN	ILE	LYS	GLY	HIS	GLY	LYS	VAL	VAL	ALA	ALA	LEU	ILE	GLU	ALA	ASN	HIS	ILE	
c	SER	HIS	GLY	SER	ALA	GLN	ILE	LYS	GLY	HIS	GLY	LYS	VAL	VAL	ALA	ALA	LEU	ILE	GLU	ALA	ASN	HIS	ILE	
d	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97
a	GAU	GAC	AUC	GCG	ACC	CUC	UCC	AAG	CUC	AGC	GAC	CAU	GCC	CAC	AAG	CUC	CGC	GUG	GAC	CCU	GUC	AAC	UUC	
b	ASP	ASP	ILE	ALA	GLY	THR	LEU	SER	LYS	LEU	SER	ASP	LEU	HIS	ALA	HIS	LYS	LEU	ARG	VAL	ASP	PRO	VAL	ASN
c	ASP	ASP	ILE	ALA	GLY	THR	LEU	SER	LYS	LEU	SER	ASP	LEU	HIS	ALA	HIS	LYS	LEU	ARG	VAL	ASP	PRO	VAL	ASN
e	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122
a	AAA	CUC	GUC	GCG	CAA	UCC	UUC	CUG	GUG	GUG	GUC	GCC	AUC	CAC	CAU	GCC	CCU	GCC	CCU	ACC	CCG	GAG	GAU	CAU
b	LYS	LEU	LEU	GLY	GLN	CYS	PHE	LEU	VAL	VAL	VAL	ALA	ILE	HIS	HIS	PRO	ALA	ALA	LEU	THR	PRO	GLU	VAL	
c	LYS	LEU	LEU	GLY	GLN	CYS	PHE	LEU	VAL	VAL	VAL	ALA	ILE	HIS	HIS	PRO	ALA	ALA	LEU	THR	PRO	GLU	VAL	
e	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141						
a	UCC	CUG	GAC	AGC	UUC	UUG	UCC	GCG	GGC	GUC	GGC	ACU	GUG	CUC	ACC	GCC	AAG	UAC	CGU	UAA	GAC	GCG	CCG	GUU
b	SER	LEU	ASP	LYS	PHE	LEU	CYS	ALA	VAL	GLY	THR	VAL	LEU	THR	ALA	LYS	TYR	ARG						
c	SER	LEU	ASP	LYS	PHE	LEU	CYS	ALA	VAL	GLY	THR	VAL	LEU	THR	ALA	LYS	TYR	ARG						
e	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	
a	GGC	CAAC	CCCAU	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	
b	GGC	CAAC	CCCAU	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	
c	GGC	CAAC	CCCAU	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	CCG	

Fig. 3 a, Nucleotide sequence of the structural globin gene cloned in p α^A HG3 (see Fig. 1); b, amino acid (aa) sequence of chicken α^A -globin chain; c, amino acid sequence of the major α -chain component of pheasant.

sequence derived from globin mRNA of phenylhydrazine-induced anaemic chicken does not code for a novel type of globin chain but for the adult α^A -globin chain that is expressed in normal chicken.

Since submission of this work, others²¹ have reported that they were unable to detect an anaemic shock α -globin gene by investigation of the genomic α -globin gene family in adult chicken.

We thank Professor H. Tiedemann for his continuous interest and generous support, and acknowledge the technical assistance of Miss E. Müller.

Received 22 October 1981; accepted 6 January 1982.

- Nienhuis, A. W. & Anderson, W. F. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2184–2188 (1972).
- Haynes, J. R. *et al. J. biol. Chem.* **255**, 6355–6367 (1980).
- Dayhoff, M. O. & Eck, R. V. *Atlas of Protein Sequence and Structure*, 154–158 (The National Biomedical Research Foundation, Maryland, 1968).
- Nienhuis, A. W. & Stamatoiyannopoulos, G. *Cell* **15**, 307–315 (1978).
- Cummings, I. W., Liu, A. Y. & Salser, W. A. *Nature* **276**, 418–419 (1978).
- Brown, J. L. & Ingram, V. M. *J. biol. Chem.* **249**, 3960–3972 (1974).
- Knöchel, W., Lange, D. & Grundmann, U. *Molec. Biol. Rep.* (in the press).
- Deacon, N. J., Shine, J. & Naora, H. *Nucleic Acids Res.* **8**, 1187–1199 (1980).
- Richards, R. I. & Wells, J. R. E. *J. biol. Chem.* **255**, 9306–9311 (1980).
- Liu, A. Y. & Salser, W. *Gene* **13**, 409–415 (1981).
- Chapman, B. S., Tobin, A. J. & Hood, L. E. *J. biol. Chem.* **255**, 9051–9059 (1980).
- Takei, H., Ota, Y., Wu, K. C., Kiyohara, T. & Matsuda, G. *J. Biochem.* **77**, 1345–1347 (1975).
- Matsuda, G., Takei, H., Wu, K. C. & Shiozawa, T. *Int. J. Protein Res.* **3**, 173–174 (1971).
- Moss, B. A. & Thompson, E. O. P. *Aust. J. biol. Sci.* **22**, 1455–1471 (1969).
- Oberthür, W., Braunitzer, G. & Kalas, S. *Hoppe-Seyler's Z. physiol. Chem.* **362**, 1101–1112 (1981).
- Alter, B. P., Goff, S. C., Efremov, G. D., Gravelly, M. E. & Huisman, T. H. J. *Br. J. Haemat.* **44**, 527–534 (1980).
- Matsuda, G., Maita, T., Mizuno, K. & Ota, H. *Nature new Biol.* **244**, 244 (1973).
- Jauregui-Adell, J. & Marti, J. *Analyt. Biochem.* **69**, 468–473 (1975).
- Zimmermann, C. L. & Pisano, J. *Meth. Enzym.* **47**, 45–51 (1977).
- Engel, J. D. & Dodgson, J. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2596–2600 (1980).
- Dodgson, J. B., McCune, K. C., Rusling, D. J., Krust, A. & Engel, J. D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5998–6002 (1981).

Univalent antibodies kill tumour cells *in vitro* and *in vivo*

M. J. Glennie* & G. T. Stevenson

Lymphoma Research Unit, Tenovus Laboratory, Southampton General Hospital, Southampton SO9 4XY, UK

Antibody molecules are bivalent, or less often multivalent, with each antibody site within a single molecule having the same specificity. Bivalency must enhance the tenacity of antibody attachment to cell surfaces, as dissociation will require simultaneous release at both sites. However, the bivalency of the antibody sometimes induces a target cell to undergo antigenic modulation^{1–3}, thereby offering the cell a means of evading complement and the various effector cells recruited by the antibody. We have investigated the attack by univalent antibodies, which, despite removal of one antibody site, retain their Fc zones and hence their ability to recruit the killing agents, on neoplastic B lymphocytes of the guinea pig L₂C line. Rabbit antibodies raised against surface immunoglobulin of these cells were partially digested with papain to yield the univalent Fab/c derivatives^{4,5}. We report here that these derivatives showed enhanced cell killing both *in vitro* and *in vivo*, and that this enhancement appeared to derive from avoiding antigenic modulation.

The L₂C lymphocytic leukaemia arose spontaneously around the year 1950 in a female guinea pig of the inbred strain 2, and has since been maintained by passaging *in vivo*; the line maintained in our laboratory was obtained in 1972 from Drs I. Green and E. Shevach (NIH). Blood from near-terminal animals was obtained by cardiac puncture and yielded L₂C preparations⁶ of >95% viability with <2% contaminating leukocytes. The cells are lymphoblastic and exhibit surface immunoglobulin M (IgM), ~50,000 molecules per cell, of light chain class λ ⁷.

* Present address: ARC Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, UK.

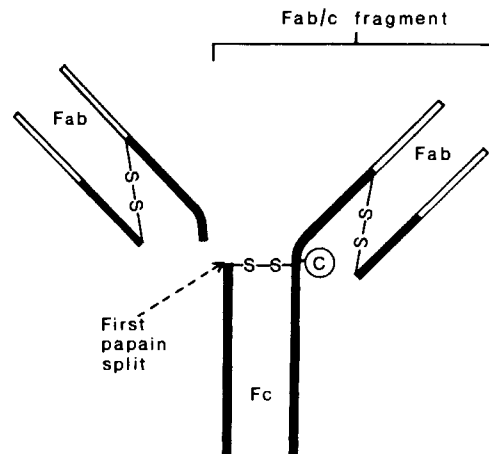


Fig. 1 Preparation of Fab/c from rabbit IgG antibody. The A12 allotypic variant of IgG¹⁷, apparently more common than the allelic A11, possesses a threonine residue on each heavy chain immediately N-terminal to the hinge-region disulphide. An oligosaccharide unit © is attached to the threonine, but on one chain only, rendering this chain relatively resistant to enzymatic cleavage in the vicinity⁸. The chain lacking the oligosaccharide unit is cleaved as shown by brief exposure to papain, yielding the Fab/c fragment.

Here we used two IgG antibodies raised in rabbits to study attack on L₂C cells. Antibody reacting with constant determinants on the λ chain of L₂C surface immunoglobulin (anti- λ) was used for studies *in vitro*; the antibody was raised against serum F(ab')₂ $\gamma\lambda$ and prepared as pure IgG antibody by elution from λ chains coupled to immunosorbent⁸. Antibody specific for idiotype determinants of the surface immunoglobulin (anti-Id) was used for studies *in vitro* and, because it is effectively tumour specific, for therapy *in vivo*⁹. It was raised against Fab μ isolated from the surface IgM⁹, rendered specific for idiotype determinants by absorption with guinea pig normal immunoglobulin, and used in the form of total IgG isolated from the absorbed antiserum.

The Fab/c derivatives of anti- λ and anti-Id, in which the antibody molecule is rendered univalent by removal of just one Fab arm^{4,5} (Fig. 1), were prepared by digesting IgG antibody with thiol-free papain (0.8 mg ml⁻¹) at 37°C for 4 h, pH 7.4. The products of digestion were fractionated by recycling chromatography on Ultrogel AcA44. To remove any contaminating F(ab')₂, the Fab/c was purified further by elution from an anti-Fc immunosorbent. In subsequent work we have used more limited exposure to papain (0.01 mg ml⁻¹ at 37°C for 30 min) in the presence of dithiothreitol (1 mM), and have been able to omit the immunosorbent step; after removing the thiol the hinge-region disulphide of Fab/c, possibly required for efficient activation of complement^{10,11}, can re-form by oxidation. Fab/c emerged from AcA44 columns in a zone corresponding to a molecular weight of ~100,000, and was obtained with a yield of 10–20% of the starting IgG. It was shown by precipitin analysis on Ouchterlony plates to possess both Fab and Fc determinants; and by gel chromatography of reduced material in 1 M propionic acid to comprise one light chain, one heavy chain and one-half of a heavy chain.

Previous work³ has shown that both anti- λ and anti-Id induce rapid antigenic modulation *in vitro*: exposure of the L₂C cells to antibody at 37°C for 2 min or longer confers resistance to lysis when they are subsequently exposed to antibody plus complement. In contrast to an earlier report describing the mouse TL system¹², bivalency of the antibody was found to be essential for induction of this modulation. It seems therefore that the redistribution of cell surface immunoglobulin induced when it is tethered by bivalent antibody—leading to patching, capping and endocytosis of the antigen–antibody complexes¹³—underlies the modulation of this antigen, although the extent of redistribution required is not clear.

Fab/c derivatives of both anti- λ and anti-Id were shown by indirect immunofluorescence to bind to the surfaces of L₂C cells. However, in contrast to the parent antibodies, there was no tendency for the Fab/c to undergo redistribution and endocytosis when attached to viable cells incubated at 37 °C for 30 min: subsequent staining by fluorescein-labelled sheep anti-rabbit IgG at 0 °C showed the Fab/c to be distributed uniformly over the cell surfaces.

The Fab/c derivatives retained the ability to invoke complement-mediated lysis of target cells, whether by guinea pig or rabbit complement. Moreover they showed no tendency to induce antigenic modulation. Figure 2a, b shows the levels of cytotoxicity induced by IgG and Fab/c of anti- λ specificity for a range of concentrations. Preincubation at 37 °C with the IgG, but not with the Fab/c, induced modulation. Note also that Fab/c induced higher levels of cytotoxicity than IgG even in experiments lacking the modulating preincubation. This difference was more marked when guinea pig complement was used, and is consistent with other evidence³ that some modulation can occur when cells are confronted simultaneously by IgG antibody and complement at 37 °C.

The findings using IgG and Fab/c from anti-Id (Fig. 2c, d) were similar. However, note that modulation by the IgG for rabbit complement was less effective than in the case of anti- λ ; that IgG antibody titred out further for rabbit complement than did Fab/c; and that the superiority of Fab/c over IgG for lysis by guinea pig complement was again apparent.

The therapeutic efficacies of IgG and Fab/c from anti-Id serum were assessed by inoculating strain 2 guinea pigs with L₂C

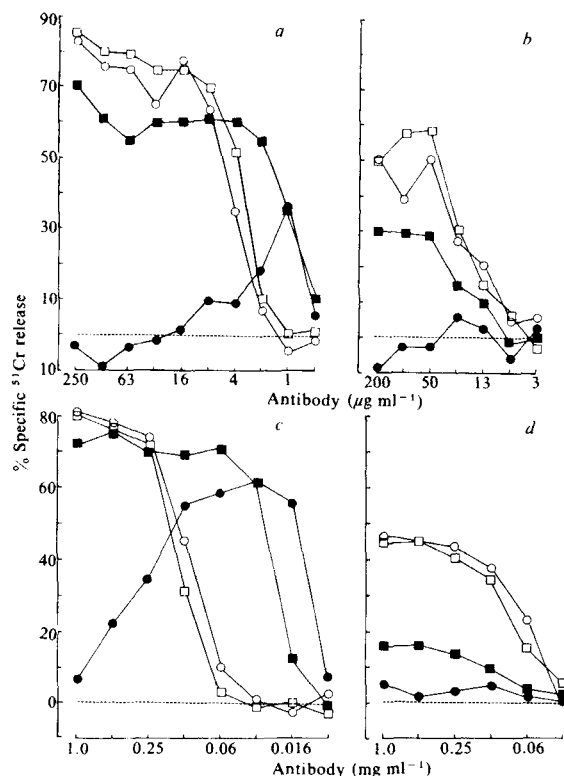


Fig. 2 Complement lysis of L₂C cells invoked by Fab/c (○, □) or IgG (●, ■) from purified anti- λ (a, b) or anti-Id serum (c, d). Complement was from rabbit (a, c) or strain 2 guinea pig (b, d). First the cells were labelled with ⁵¹Cr and washed. To permit modulation they were incubated at 5 × 10⁵ ml⁻¹ with antibody derivatives at the indicated concentrations at 37 °C for 15 min (○, ●); control incubations were at 0 °C (□, ■). All suspensions were placed in an ice bath; 1.5 vol of a 1:2 dilution of fresh serum was added as a source of complement, and the temperature was raised to 37 °C for 30 min to promote complement activity. Finally the suspensions were chilled and the supernatants separated for assay of released ⁵¹Cr. Methods are described in detail in ref. 19.

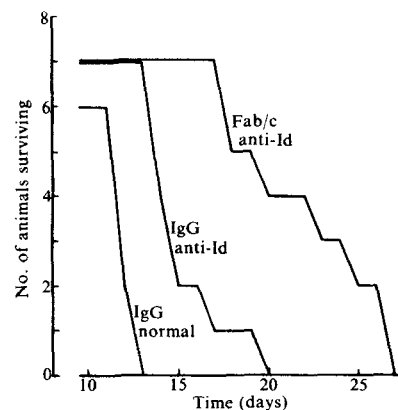


Fig. 3 Immunotherapy with Fab/c or IgG from anti-Id serum. Groups of guinea pigs were inoculated intraperitoneally (i.p.) with 10⁵ L₂C cells (day 0). After 24 h they were injected i.p. with 10 mg IgG from normal sheep serum (6 animals), 10 mg IgG from anti-Id serum (7 animals), or 10 mg Fab/c derivative of the anti-Id IgG (7 animals). Animals treated with IgG (anti-Id) survived longer than those treated with IgG (normal), as did those treated with Fab/c (anti-Id) compared with IgG (anti-Id)-treated animals ($P < 0.005$, χ^2 test²⁰).

leukaemia, then injecting them once with the test substance 24 h later, and observing survival (Fig. 3). Our tumour line has a doubling time *in vivo* of ~19 h, thus a simple kinetic model suggests that 50% killing of the tumour will prolong life by this time period. The complexity of events occurring *in vivo* introduces considerable uncertainty, but evidently the greater the cell kill from a single therapeutic event, the greater the subsequent average survival, with the duration of survival permitting some tentative inferences about the extent of killing. By increasing average survival by 10 days Fab/c was much more effective than the parent IgG (3.2 days), and in fact much more effective than any of our previous attempts at antibody therapy.

We have not yet defined the relative involvements of the classical and alternative complement pathways in the cell lysis invoked by IgG and Fab/c. Activation of the classical pathway by Fab/c would be further evidence against the possibility that steric changes in the IgG molecule consequent on both Fab arms being engaged are necessary for the binding and activation of complement component 1 (C1) (reviewed in ref. 14). The alternative pathway acts as an amplifier after activation of the classical pathway, and apparently acts independently in the lysis of certain virus-infected cells by antibody and complement¹⁵. The fact that bivalent antibody was required in the latter studies makes it unlikely that Fab/c invokes a similar mechanism.

The clear superiority of Fab/c over IgG antibody in invoking lysis by homologous (guinea pig) complement confirms the importance of antibody bivalency in permitting mammalian cells to escape from antibody-mediated destruction. Eluding this escape mechanism appears to offer a surprisingly simple means of enhancing the attack of antibodies on tumour cells. We are now evaluating Fab/c and similar derivatives of anti-idiotypic antibody in the treatment of human chronic lymphocytic leukaemia, where the cells show some susceptibility to cytotoxic attack by anti-idiotypic IgG¹⁶.

This work was supported by the Cancer Research Campaign, the MRC and Tenovus.

Received 7 September 1981; accepted 15 January 1982.

- Boyse, E. A. & Old, L. J. *A. Rev. Genet.* **3**, 269–290 (1969).
- Stackpole, C. W., Jacobson, J. B. & Lardis, M. P. *J. exp. Med.* **140**, 939–953 (1974).
- Gordon, J. & Stevenson, G. T. *Immunology* **42**, 13–17 (1981).
- Goodman, J. W. *Biochemistry* **4**, 2350–2357 (1965).
- Michaelsen, T. E. & Natvig, J. B. *Scand. J. Immun.* **1255**–1268 (1972).
- Stevenson, G. T., Eady, R. P., Hough, D. W., Jurd, R. D. & Stevenson, F. K. *Immunology* **28**, 807–820 (1975).

7. Hough, D. W., Chapple, J. C., Stevenson, F. K. & Stevenson, G. T. *Immunology* **34**, 889–899 (1978).
8. Stevenson, F. K., Elliott, E. V. & Stevenson, G. T. *Immunology* **32**, 549–557 (1977).
9. Stevenson, G. T., Elliott, E. V. & Stevenson, F. K. *Fedn Proc.* **36**, 2268–2271 (1977).
10. Press, E. M. *Biochem. J.* **149**, 285–288 (1975).
11. Isenman, D. E., Dorrington, K. J. & Painter, R. H. *J. Immun.* **114**, 1726–1729 (1975).
12. Lamm, M. E., Boyse, E. A., Old, L. J., Lisowska-Bernstein, B. & Stockert, E. *J. Immun.* **101**, 99–103 (1968).
13. Taylor, R. B., Duffus, P. H., Raff, M. C. & de Petria, S. *Nature new Biol.* **233**, 225–227 (1971).
14. Metzger, H. *Contemp. Topics molec. Immun.* **8**, 119–152 (1978).
15. Sissons, J. G. P. et al. *J. exp. Med.* **150**, 445–454 (1979).
16. Hamblin, T. J., Abdul-Ahad, A. K., Gordon, J., Stevenson, F. K. & Stevenson, G. T. *Br. J. Cancer* **42**, 495–502 (1980).
17. Prahl, J. W., Mandy, W. J. & Todd, C. W. *Biochemistry* **8**, 4935–4940 (1970).
18. Fanger, M. W. & Smyth, D. G. *Biochem. J.* **127**, 767–774 (1972).
19. Gordon, J., Robinson, D. S. F. & Stevenson, G. T. *Immunology* **42**, 7–12 (1981).
20. Peto, R. *Br. J. Cancer* **29**, 101–105 (1974).

Homopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S₁ nuclease

Christopher C. Hentschel*

Institut für Molekularbiologie II der Universität Zürich,
Hönggerberg, 8093 Zürich, Switzerland

The concept of specific functional elements in genes comprising localized or transient structural discontinuities, such as Z-DNA or cruciforms, is becoming increasingly plausible^{1–6}. Cruciforms were first detected as supercoil-dependent single-strand nuclease (S₁)-sensitive sites characterized by hyphenated inverted repeat DNA sequences^{2,3}. Here I describe examples of a novel type of S₁ nuclease site found in a sea urchin histone gene repeat and characterized by homopolymer sequences. Endonucleolytic cleavage within these sequences does not depend on supercoiling, is highly sensitive to the salt concentration and shows a reproducible pattern of maxima and minima. This type of S₁ sensitivity may reflect the potential for out-of-register DNA slippage in such sequences and I speculate that related slippage events *in vivo* could lead to their acting as foci for recombinational events.

Recombinant plasmids containing either one copy of the 6.25 kilobase (kb) *Psammechinus miliaris* histone gene repeat h22 (refs 7, 8) inserted in the *Hind*III site of pBR322 (ref. 9; pBRh22/3), or a head-to-tail dimer of the repeat (pBRh22/4) similarly inserted, were digested with low levels of the single-strand-specific endonuclease¹⁰ S₁ in conditions which linearized supercoiled pBR322 alone^{2,3}. Surprisingly these plasmids were not only linearized but yielded discrete excision products indicating the presence of at least one supercoil-independent S₁ site (Fig. 1 *a, b*). In addition, the pattern of excision products bore no relation to A + T-rich stretches or inverted repeats in the DNA sequence. Previous studies have only detected specific S₁ sites in 'duplex' DNA, either at inverted repeats in supercoiled molecules^{2–4}, or in conditions which partially denature A + T-rich regions (for reviews, see refs 5, 6).

At the S₁ nuclease concentration used (see Fig. 1 legend), there was little background smear from random cleavage at A + T-rich regions but only 'partials' of the specific cleavage reactions were produced. It is nevertheless clear that whether one or two copies of the histone gene repeat is present, a 1.6 kb fragment is always excised. A 6.25 kb repeat length fragment and a 4.6 kb fragment (which together with the 1.65 kb fragment sums to the repeat length) appear only with the dimer insert (Fig. 1 *a*). These results are consistent with the presence of only two major S₁ sites in the histone repeat sequence, 1.65 kb apart. Furthermore, as no excision products resulted from identical S₁ digestion of a 'mutant' plasmid lacking the *Hpa*I fragment

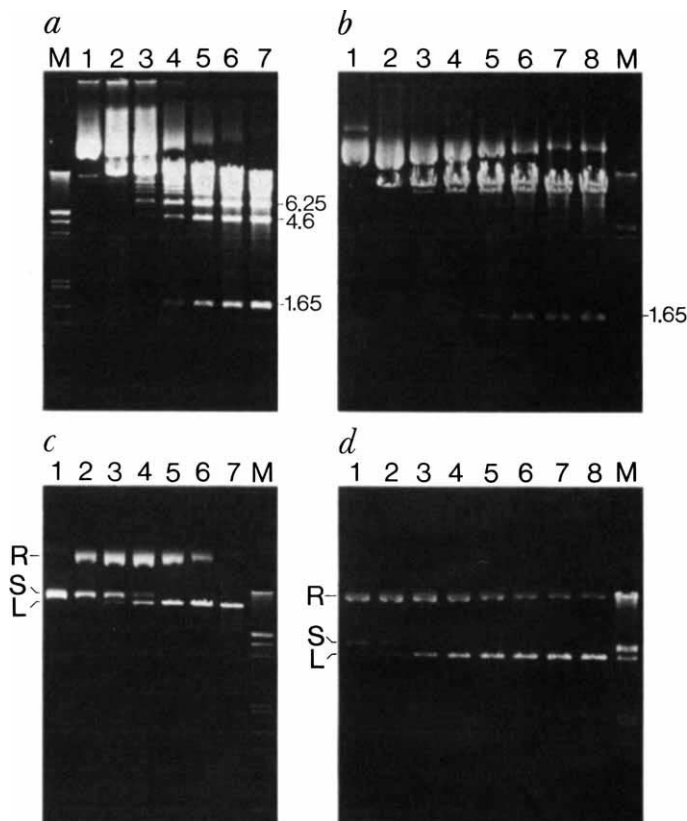


Fig. 1 Effect of S₁ nuclease on four plasmids analysed by 1% agarose gel electrophoresis; *a*, pBRh22/4, a recombinant containing a head-to-tail dimer of the 6.25 kb *Hind*III-cut histone repeat h22⁷, inserted into the *Hind*III site of pBR322⁹; *b*, pBRh22/3, identical to pBRh22/4 but with only one histone repeat; *c*, a derivative of pBRh22/3 lacking the *Hpa*I fragment indicated by a thick line in Fig. 2*c*; *d*, pBR322. Lanes marked M are λ /*Hind*III/*Eco*RI size markers²⁶. Other lanes are time points of S₁ nuclease digestion (lanes 1–8, respectively, 0, 2, 4, 8, 16, 32, 64 and 132 min). S₁ nuclease, prepared according to Vogt¹⁰, was used at a concentration of 50 Vogt units per ml in 30 mM NaAc pH 4.6, 30 mM NaCl, 1 mM ZnCl₂ at 37 °C. Electrophoresis buffer was 100 mM Na/glycine pH 8.5 containing 0.5 μ g ml⁻¹ ethidium bromide for UV photography²⁷. For the smaller plasmids (*c, d*), the supercoiled (S), relaxed or nicked (R) and linear (L) forms are indicated. Sizes are in kb.

spanning the H2A coding region and the H2A–H1 spacer (Fig. 1*c* and map in Fig. 2), one of the sites was tentatively located within this region of the histone gene repeat. The mutant plasmid was linearized at the second histone repeat site faster than pBR322 alone (Fig. 1*d*; mapping data not shown).

To locate exactly the two major S₁ sites on h22, S₁-cut pBRh22/3 and pBRh22/4 were secondarily restricted with several enzymes. These double digests mapped the S₁ sites in the spacers flanking the H1 gene. As anticipated, one of the sites lay within the *Hpa*I fragment and the other 1.65 kb downstream (see map in Fig. 2); this was confirmed by purifying the 1.65 kb S₁ excision product and restricting it with either *Bam*HI, yielding 900 and 750 base pair (bp) fragments, or *Bgl*II, yielding 850 and 800 bp fragments (Fig. 2*a*). These fragments are less clearly defined on the gel than the marker restriction fragments, suggesting that the S₁ cut ends are heterogeneous. Such heterogeneity is perhaps not surprising because the DNA sequences to which they map comprise the simple repeat motifs 5'(CA)10(CT)22 [site S₁ (*a*) in Fig. 2; P. Bucher, unpublished results] and 5'(GA)16 [site S₁ (*b*) in Fig. 2; ref. 7]. The S₁ (*b*) site is so close to the *Hind*III site downstream of the H1 gene (Fig. 2*c*) that when the *Hind*III-excised repeat is S₁-cleaved again, cutting at this site is not easily discernible, although it does take

* Present address: Department of Molecular Genetics, Celltech Limited, 250 Bath Road, Slough SL1 4DY, UK.

place at a significant level (data not shown). However, partial cleavage at the $S_1(a)$ site within this linear molecule is easily demonstrable and results in the production of 1.75 and 4.5 kb fragments (Fig. 2b). Thus cutting at either $S_1(a)$ or (b) does not strictly depend on supercoiling, although an effect on the cutting rate has not been excluded. Moreover the larger $S_1(a)$ site seems to be inherently more S_1 sensitive, at least in linear molecules (Fig. 2b and data not shown).

The observation that the homopolymers flanking the H1 gene constitute novel S_1 -sensitive sites prompted an examination of the possible mechanism involved. Because the S_1 activity is independent of DNA supercoiling, models based on the formation of unpaired regions in structures such as cruciforms²⁻⁴, stabilized by torsional stresses, can be excluded. Moreover, unlike supercoil-dependent S_1 sites², the S_1 cleavage at the homopolymers is highly sensitive to salt concentration: excision of the 1.65 kb fragment is strongly inhibited at 120 mM Na^+ , close to the optimum salt conditions for S_1 activity, and is all but abolished at 300 mM, a concentration at which most of the intrinsic S_1 activity still remains^{10,11} (Fig. 3). As duplex stability is directly proportional to the logarithm of total salt concentration¹², this result strongly implies that the S_1 sensitivity of the homopolymers reflects a pre-melting phenomenon specifically associated with such simple repeated sequences¹³.

A possible explanation, based on studies with pure homopolymer preparations¹³, is out-of-register DNA slippage with the production of single-strand loops flanking the slipped segment. Such structures would be expected to exist transiently. This notion is supported by examination of the distribution of nucleolytic cleavage within the homopolymer sequences. For Fig. 4a, S_1 -cut, ³²P-end-labelled pBRh22/4 was secondarily cleaved either with *Bam*HI (as a control to establish that label was incorporated in the anticipated 900 and 750 base fragments adjacent to the two S_1 sites; lane 1), or with the frequent cutting enzymes *Hae*III (lane 2), *Alu*I (lane 3) or *Taq*I (lane 4). The fragments were displayed on a sequencing gel and autoradiographic examination of the gel reveals the double digestion products expected from the S_1 site map (Fig. 2c) and the restriction map¹. Note that the gel fails to resolve the terminal heterogeneity of the S_1 -cleaved ends of the larger fragment but

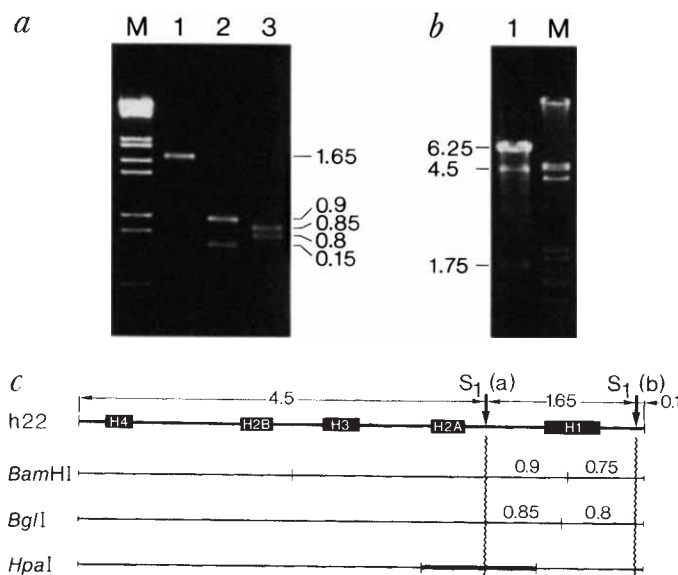


Fig. 2 Restriction enzyme mapping of S_1 cleavage sites. *a*: Lane 1, purified 1.65 kb S_1 excision fragment; 2, *Bam*HI restriction products of the 1.65 kb fragment; 3, *Bgl*II restriction products of the 1.65 kb fragment. *b*: Partial cleavage of *Hind*III repeat of h22 with S_1 nuclease using the conditions described in Fig. 1 legend with digestion for 1 h. The electrophoretic analyses were as for Fig. 1 except that for *a*, 2% agarose was used. *c* is a map of the h22 repeat showing the deduced S_1 site map relative to the restriction map¹.

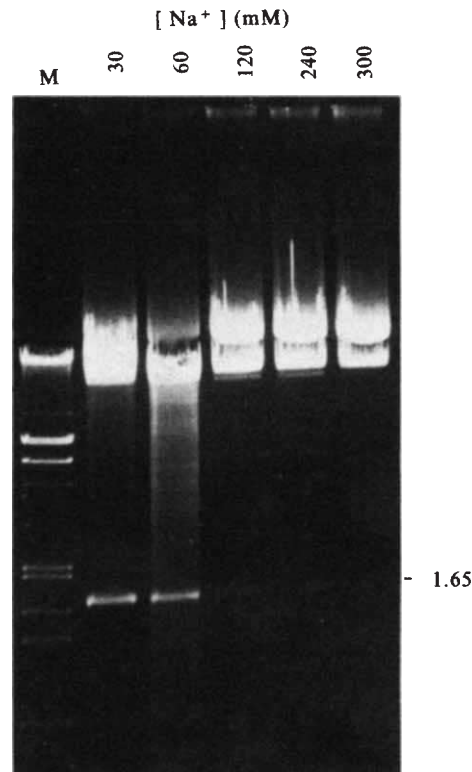


Fig. 3 Salt sensitivity of S_1 cleavage. pBh22/3 was digested with S_1 for 32 min with conditions and electrophoretic analysis as for Fig. 1 except that the NaCl concentration was adjusted to bring the total Na^+ concentration to the values indicated. The degree of S_1 cleavage can be judged by the production of the 1.65 kb excision product.

the smaller (≤ 200 bp) fragments are clearly resolved in a manner reflecting the nucleolytic cleavage pattern within the homopolymers. Thus for the $S_1(b)$ site, this pattern is bimodal and cutting is largely restricted to within a stretch of ~ 20 nucleotides (bracketed in Fig. 4). Figure 4c shows the sequence around $S_1(b)$ with the relevant restriction sites and the deduced pattern of S_1 cleavage within the 5'(GA)16 motif. Thus, for example, *Taq*I restriction (Fig. 4a, lane 4) produces only one group of small terminal fragments (65–85 bases long), which can be seen to map from the *Taq*I site (■; Fig. 4c); similarly the *Alu*I site (●) maps the 95–115 base fragments. In fact all the bands shown in Fig. 4a can be mapped from an appropriate restriction site to either $S_1(b)$ or $S_1(a)$, although in the latter case the fragments are generally larger due to the paucity of restriction sites in the H2A–H1 spacer¹ (sequence not shown).

A bimodal cleavage pattern would be expected from a single predominant type of DNA slip as this creates two single-strand loops adjacent to the slip, each a potential S_1 target. This seems to be compatible with the $S_1(b)$ pattern where the site comprises a simple homopolymer. The cleavage fragments mapping to the $S_1(a)$ site show a more complex pattern, approximately trimodal with a predominant central cleavage region covering ~ 40 nucleotides (flanked by a wavy line in Fig. 4). This greater length and complexity may be a time-averaged reflection of the many slipping modes possible within a long dual homopolymer such as the $S_1(a)$ site (5'(CA)10(CT)22). Figure 4b illustrates an alternative 'fine mapping' strategy. The purified 1.65 kb S_1 -excised fragment was ³²P-end-labelled¹⁴ and the derived *Bam*HI fragments (Fig. 2a) subjected to total chemical cleavage at either pyrimidine (lanes 1, 3) or purine (lanes 2, 4) nucleotides. Since the 5'-end-label lies almost exclusively in C-T runs for both *Bam*HI fragments, pyrimidine cleavage yielded no oligonucleotides when analysed on a 20% acrylamide gel (Fig. 4b, lanes 1, 3). In contrast, purinolysis mapped the S_1 cleavage pattern from neighbouring purine nucleotides (Fig. 4b, lanes

2, 4). Figure 4b reproduces almost exactly the complex multimodal pattern previously seen for the $S_1(a)$ site extending over about 40 nucleotides, and the bimodal pattern for $S_1(b)$ extending over about 20 nucleotides (cf. Fig. 4a). Note also that these map identically on the sequence (the purines from which to map the $S_2(b)$ pattern are complementary to the 5'TT in Fig. 4c).

Whatever the physical basis for the S_1 nuclease sensitivity of the homocopolymers flanking the H1 gene of the h22 histone gene repeat, it is tempting to speculate about the possible biological significance of the localized DNA unpairing implied by these results. Homocopolymer sequences are found not only

in the h22 repeat, but also adjacent to histone genes in several species (not necessarily the H1 gene)⁸, the sequence 5'(CA)31 has been found ~300 bases downstream of the coding sequence of two immunoglobulin genes¹⁵, and 5'(TG)17 has recently been found in an evolutionarily conserved repeat family in, for example, the 'spacer' separating the human δ - and β -globin genes¹⁶.

The homocopolymer sequences of the histone genes are located in spacer regions not known to have a direct effect on the expression of the genes⁸, but which have been strongly implicated in evolutionary behaviour¹⁷. Although histone genes are usually clustered or tandemly repeated, the copy number, order and polarity of the genes vary widely between species and, in some cases, even within a single genome. Where the copy number is high, as in sea urchins or *Drosophila*, tandem repetition is the norm but escaped 'orphan' histone genes arise with a high frequency¹⁸. Subtypes of H1 genes arise at a particularly high frequency (see reviews in refs 8, 17, 19, 20, 13–15 and refs therein). All these facts suggest that mechanisms exist which allow occasional histone genes (and in particular H1 genes) to transpose, replicate and diverge independently. Could *in vivo* excision events at homocopolymers be involved in such mechanisms?

The rectification mechanism maintaining the homogeneity of tandemly repeated histone genes may also involve the homocopolymers because they could provide foci for unequal crossover or gene conversion events¹⁷. Although the exact nature of such recombinations is unknown, they would probably be facilitated by a potential for localized unpairing in these sequences. Moreover, the possibility that such unpairing reflects DNA slippage and that similar slippage may occur *in vivo* is supported by the fact that several biological phenomena are best explained by this mechanism: frequent frameshift mutations in the T4 lysozyme gene²¹, mutational hotspots in the *LacI* gene of *Escherichia coli*²² and small deletions in human globin genes²³. In addition, other specific recombinational events, such as globin gene conversion²⁴ or immunoglobulin class switching²⁵, are also thought to occur at simple repeated sequences. All these events could be related to a fundamental tendency for DNA slippage at such sequences.

I thank Professor Max Birnstiel for initially suggesting the DNA slippage mechanism, Bill Folk and Mary Bendig for discussion and critical reading of the manuscript, and Fritz Ochsenbein, Bonnie Bryan and Jackie Hardwick for help in its preparation. This work was supported by a grant from the State of Zurich and by Swiss NRF grant 3/257.077.

Received 2 October 1981; accepted 8 January 1982.

1. Nordheim, A. *et al.* *Nature* **294**, 417–422.
2. Lilley, D. M. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6468–6472 (1980).
3. Panayotatos, N. & Wells, R. D. *Nature* **289**, 466–470 (1981).
4. Lilley, D. M. J. *Nucleic Acids Res.* **9**, 1271–1289 (1981).
5. Wells, R. D. *et al.* *Prog. Nucleic Acid Res. molec. Biol.* **24**, 167–267 (1980).
6. Cantor, C. R. *Cell* **25**, 293–295 (1981).
7. Schaffner, W. *et al.* *Cell* **14**, 655–671 (1978).
8. Hentschel, C. C. & Birnstiel, M. L. *Cell* **25**, 301–313 (1981).
9. Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2727 (1978).
10. Vogt, V. *Eur. J. Biochem.* **33**, 192–200 (1973).
11. Sutton, W. D. *Biochim. biophys. Acta* **240**, 522 (1971).
12. Schildkraut, C. & Lifson, S. *Biopolymers* **3**, 195 (1968).
13. Cantor, C. R. & Schimmel, P. R. in *Biophysical Chemistry* Vol. 3, Ch. 22 (Freeman, San Francisco).
14. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
15. Nishioka, Z. & Leder, P. *J. biol. Chem.* **255**, 3691–3694 (1980).
16. Miesfeld, R., Krystal, M. & Arnheim, N. *Nucleic Acids Res.* **9**, 5931–5947 (1981).
17. Kedes, L. H. *Ann. Rev. Biochem.* **48**, 837–870 (1979).
18. Childs, G., Maxson, R., Cohn, R. & Kedes, L. H. *Cell* **23**, 651–663 (1981).
19. Isenberg, I. A. *Rev. Biochem.* **48**, 159–191 (1979).
20. Hohmann, P. *Int. Rev. Cytol.* **71**, 41–93 (1971).
21. Streisinger, G. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **31**, 77–84 (1966).
22. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. *J. molec. Biol.* **126**, 847–863 (1978).
23. Efstratiadis, A. *et al.* *Cell* **21**, 653–668 (1980).
24. Slightom, J. L., Blech, A. E. & Smithies, O. *Cell* **21**, 627–638 (1980).
25. Nikaido, T., Nakai, S. & Hongo, T. *Nature* **292**, 845–848 (1981).
26. Southern, E. *Meth. Enzym.* **68**, 152–176 (1979).
27. Sharp, P. A., Sugden, B. & Sambrook, J. *Biochemistry* **12**, 3055–3063 (1973).
28. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107 (1978).
29. Hofstetter, H., Kressmann, A. & Birnstiel, M. L. *Cell* **24**, 573–585 (1981).

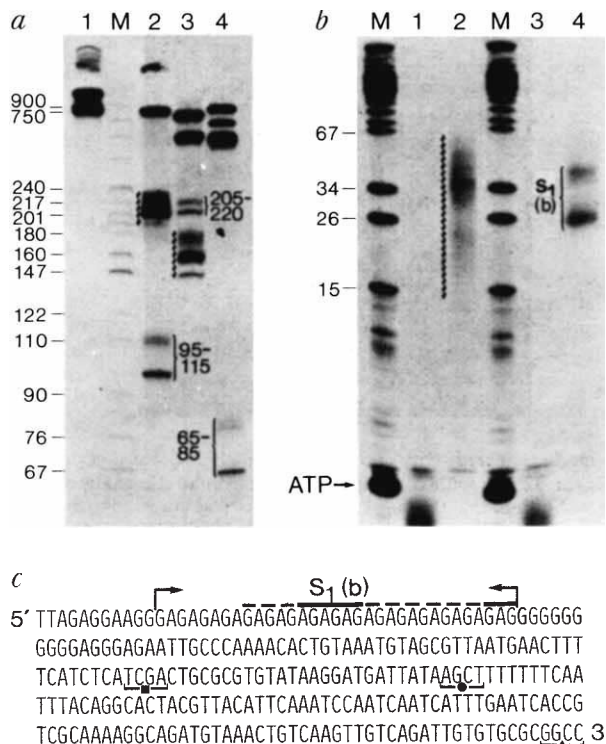


Fig. 4 Fine mapping of nucleolytic cleavage sites. *a*: Autoradiograph of 8% acrylamide/urea gel²⁸ of fragments derived by secondarily cutting S_1 -digested 5'-³²P-end-labelled pBh22/4 (ref. 14) with the following enzymes: lane 1, *Bam*HI; lane 2, *Hae*III; lane 3, *Alu*I; lane 4, *Taq*I. The size marker (in bp; M) is a *Hpa*II-cut derivative of pBR322 (ref. 29). Note that in these experiments there are eight labelled ends which would be expected to yield four predominant labelled fragments assuming that nearly all the initial S_1 cleavage occurred as indicated in Fig. 2. Thus *Bam*HI should yield fragments 750 and 900 bases long (barely resolved in lane 1) and 2.1 and 2.3 kb (not resolved; see Fig. 2c). All the radioactivity is in the expected region of the gel and the sizes were confirmed on an agarose gel (not shown); most important, however, no 'unexpected' low molecular weight fragments are present, thus the heterogeneous bands in lanes 2–4 must derive from the anticipated double digestions. The heterogeneous fragments derived from cleavage at $S_1(b)$ are shown bracketed, while those from $S_1(a)$ are flanked by a wavy line. *c* Shows the sequence for the $S_1(b)$ site in 50 base blocks with the relevant restriction sites (for *Taq*I, *Alu*I *Hae*III) and the deduced regions of maximal cleavage indicated by overlining. *b* Shows an alternative 'fine mapping' strategy. An aliquot of the same purified 1.65 kb fragment shown in Fig. 2a was 5'-³²P-end-labelled¹⁴, and the derived *Bam*HI fragments (Fig. 2a, lane 2) purified. Each fragment was divided into two aliquots which were either totally hydrolysed at purine nucleotides with formic acid/diphenylamine reagent (lanes 2, 4) or at pyrimidine nucleotides by hydrazinolysis (lanes 1, 3), and then purified as for sequencing reactions¹⁴. The purified products were analysed by autoradiography following electrophoresis on a 20% acrylamide/urea gel¹⁴. The products from the 900 bp fragment mapping the $S_1(a)$ sites are in lanes 1 and 2, while those from the 750 bp fragment mapping the $S_1(b)$ site are in lanes 3 and 4. Size marker as in *a*.

MATTERS ARISING

Magma chamber profiles from Bay of Islands ophiolite complex

CASEY and Karson¹ have presented a model for the shape and size of magma chambers underlying mid-ocean ridges, based on observations of the Bay of Islands ophiolite. They show that igneous layering in rocks of the plutonic complex is oriented up to 90° oblique to the palinspastic horizontal, and they use this, and other evidence, to infer that nucleation and growth of crystals occurred *in situ* on the chamber walls, without the need for settling of crystals.

The magma chamber model of Casey and Karson (Figs 4 and 5 of ref. 1) shows a curious, unexplained and yet crucial feature, namely that while ultramafic rocks were forming along the sides of the chamber near its base, mafic rocks were forming along the sides at higher levels. Neither the effect of pressure nor that of volatile content on crystallization of the magma can be responsible for this difference. If the ultramafic and the mafic rocks indeed formed in the same magma chamber, and if Casey and Karson are correct in saying that crystallization occurred only along the chamber walls, then the magma must have been stratified, with more mafic magma lying below less mafic magma; the 'Unit 2' rocks of the plutonic complex (interlayered mafic and ultramafic rocks) would then represent the junction between the two magma types, that is a zone of mixing.

This inference that the magma was compositionally stratified agrees with that reached by Sparks *et al.*² on the basis of the density variation in mid-ocean-ridge basalts.

C. H. DONALDSON

*Department of Geology,
St Andrews University,
Fife KY16 9ST, UK*

1. Casey, J. F. & Karson, J. A. *Nature* **292**, 295–301 (1981).
2. Sparks, R. S. J., Meyer, P. & Sigurdsson, H. *Earth planet. Sci. Lett.* **46**, 419–430 (1980).

CASEY AND KARSON REPLY—Donaldson's comment that we have inferred the existence of chemical heterogeneity along the margin of the magma chamber which formed the plutonic complex of the Bay of Islands Ophiolite is perhaps an understatement. We categorically stated that "the lack of lateral continuity of the fine-scale layering must be interpreted as the result of fairly significant changes in physico-chemical conditions on a scale approximated by the lateral continuity of the layering ($\sim \leq 50$ m)". We also stated that "sig-

nificant chemical zonation in the magma chamber" was reflected by the fact "that fine scale igneous layers (which define isochronous planes) strike obliquely across the contacts between major lithologic subdivisions of the plutonic section".

Donaldson has also suggested that this interpretation of chemical zonation is consistent with the work of Sparks *et al.*¹ (and presumably Huppert and Sparks^{2,3}). He has further postulated that the transition zone ('Unit 2') of the plutonic complex can be interpreted as a mixing zone between two distinct magma types. Certainly, Sparks and his co-workers have clearly shown the weaknesses inherent in the arguments which lead to the emphatic statements that mafic magma chambers cannot become chemically zoned because of continuous convective overturn. Based on the extent of lateral continuity of layering throughout the plutonic complex and other factors treated in the original paper, however, we do not believe that one can simply categorize two magma types. The overall chemical and mineralogical zonation of the crystallization products within the plutonic complex seems to span the entire thickness of the plutonic section. Previous geochemical studies^{4–6} and unpublished data collected by us indicate that there is no sharp break in mineral compositions identified at the transition zone. These data seem to indicate that at any single time, the margin of the chamber was in contact with magma of numerous and highly variable compositions. The compositional range seems to depend on stratigraphic height within the plutonic section. The transition zone simply marks the first appearance of plagioclase as a major phase within the plutonic section.

Detailed mineral chemistry studies by our group will attempt to characterize the scale and degree of chemical variations along a single magma chamber profile. Until such time, however, it is premature to suggest a simple model having just two magma types and a single mixing zone. The chamber could be multiply stratified on a fine scale and consist of multiple mixing zones that need not have a steady-state position within the chamber. The dimension of the chamber might also be conducive to local chaotic and laterally discontinuous chemical stratification. These possibilities have not yet been tested by Sparks and co-workers because they have initiated their studies using simple two liquid models. It is also not clear that pressure and volatile content of the magma should be overlooked or discarded as a possible cause for at least part of the chemical and mineralogical zonation until further work, especially trace element and isotopic studies, has been completed. At present there is little evidence to substantiate Donaldson's simplistic model of just two magma types;

field and geochemical evidence seems to favour a more complex multiple stratified model.

J. F. CASEY

*Department of Geology,
University of Houston,
Central Campus,
Houston, Texas 77004, USA*

J. A. KARSON

*Department of Geology and
Geophysics,
Woods Hole Oceanographic
Institution,
Woods Hole,
Massachusetts 02543, USA*

1. Sparks, R. S. J., Meyer, P. & Sigurdsson, H. *Earth planet. Sci. Lett.* **46**, 419–430 (1980).
2. Huppert, H. E. & Sparks, R. S. J. *Contr. Miner. Petrol.* **75**, 279–289 (1980).
3. Huppert, H. E. & Sparks, R. S. J. *Nature* **286**, 5768, 46–48 (1980).
4. Irvine, J. N. & Findley, T. C. *Publ. Earth Phys. Br. Dept. of Energy, Mines Resources, Can.* **42**, 97–128 (1972).
5. Church, W. R. & Riccio, L. *Can. J. Earth Sci.* **14**, 1156–1165 (1977).
6. Malpas, J. thesis, Memorial Univ. Newfoundland (1976).

K⁺ depletion and Na⁺ pump density

IT has been shown¹ that the skeletal muscles of K⁺-depleted rats lose large quantities of K⁺, gain Na⁺, and many have fewer ouabain binding sites than normal; this is in contrast to the findings of Chan and Sanslone² and Erdman, Bolte and Luderitz³ for red blood cells (RBCs) and heart muscle, respectively, of K⁺-depleted rats and in cultured human cells grown in low-K⁺ medium^{4,5}.

In our original experiments we showed⁴ that more pumps were produced in cells when the internal Na⁺ concentration ($[Na^+]_i$) was allowed to rise to 20–60 mM, with a corresponding drop in $[K^+]_i$. In recent experiments we have found that very large increases in $[Na]_i$ (and correspondingly large reductions in $[K^+]_i$) lead to a large loss of ouabain binding sites and transport capacity in HeLa cells (Table 1). These losses in ouabain binding sites are greater than can be accounted for by cessation of the normal insertion rate of new pumps ($1.5\% h^{-1}$)⁶ and imply that an accelerated removal of pumps is occurring.

Our results show that HeLa cells can be made to produce either a greater or lesser density of pumps than normal, apparently depending on the ion concentrations within the cells; they therefore mimic RBCs and heart or skeletal muscles of K⁺-depleted rats. We suspect that two processes are occurring: (1) as $[Na]_i$ rises, the rate of insertion of new pumps increases, and (2) if $[K]_i$ falls below a critical level, the synthesis of new pumps stops (because K⁺ is required as a co-factor by most enzymes⁷) and pumps are

Table 1 Effect of changes in extracellular K^+ on the intracellular Na^+ and K^+ concentrations and on the specific ouabain binding of HeLa cells

Expt	$[K]_o$ (mmol per 1 H_2O)	$[Na]_i$	$[K]_i$	Specific ouabain binding as % of control
Ref. 4	0.4 $\rightarrow$ 0.6	24	145	159
1	0.1 $\rightarrow$ 0.3	70	122	136
2	0.2 $\rightarrow$ 0.4	87	122	96
1	0.1 $\rightarrow$ 0.2	111	35	83
3	0.2 $\rightarrow$ 0.3	97	46	55
4	0.04 $\rightarrow$?	165	50	17

removed from the membrane (by affecting the anchoring process?).

It is possible that the mechanism controlling the density of Na^+ pumps in skeletal muscle is no different from that in other cells, but that the ion concentrations achieved by any reduction of serum $[K^+]$ differ. This could be investigated further by looking at the effects of smaller reductions of serum $[K^+]$ on the ouabain binding and $(Na^+ + K^+)ATPase$ of rat skeletal muscle.

Control cells had a $[Na]_i$ of 16 mmol l^{-1} and a $[K]_i$ of 179 mmol l^{-1} . The $[K]_o$ values are the initial and final values (where measured).

J. F. LAMB
J. F. AITON
PAT OGDEN

Department of Physiology
and Pharmacology,
Bute Medical Buildings, St Andrews,
Fife KY16 9TS, UK

- Nørsgaard, A., Kjeldsen, K. & Clausen, T. *Nature* **293**, 739–741 (1981).
- Chan, P. C. & Sanslone, W. R. *Archs Biochem. Biophys.* **134**, 48–52 (1969).
- Erdman, E., Bolte, H.-D. & Lüderitz, B. *Archs Biochem. Biophys.* **145**, 121–125 (1971).
- Boardman, L., Huett, M., Lamb, J. F., Newton, J. P. & Polson, J. M. *J. Physiol., Lond.* **241**, 771–794 (1974).
- Pollack, L. R., Tate, E. H. & Cook, J. S. *J. cell. Physiol.* **106**, 85–97 (1981).
- Lamb, J. F. & Ogden, P. Q. *Jl exp. Physiol.* **67**, 105–119 (1982).
- Baker, P. F. *Endeavour* **25**, 166 (1966).

NØRGAARD ET AL. REPLY—Our observation that K^+ -depletion in mice and rats leads to a progressive decrease in the number of 3H -ouabain binding sites in skeletal muscle¹, does not agree with the results of previous investigators, who found that K^+ depletion increases $(Na^+ + K^+)ATPase$ and the number of 3H -ouabain binding sites in erythrocytes^{2,3} as well as $(Na^+ + K^+)ATPase$ activity in guinea pig hearts^{4,5}. Our observation that erythrocytes of K^+ -depleted rats maintain normal K^+ -contents even after weeks of exposure to the low plasma K^+ values agrees with the existence of a compensatory rise in their number of $(Na^+ + K^+)ATPase$ units. The reported measurements^{4,5} of $(Na^+ + K^+)ATPase$ activity in hearts used a sediment obtained by centrifugation of homogenates, and as this may contain only a minor fraction of the total $(Na^+ + K^+)ATPase$ activity present in the starting material (for discussion, see ref. 6), the results depend on the recovery

being the same for hearts from normal and K^+ -depleted animals.

To explore this problem further, we have recently measured the binding of 3H -ouabain to rat heart ventricles *in vivo*. In the controls we found $276 \pm 8 \text{ pmol per g wet wt}$, and $157 \pm 18 \text{ pmol per g}$ in the K^+ -depleted rats ($P < 0.001$). This decrease (43%) is smaller than that found in skeletal muscle, which agrees with the observation that the loss of K^+ was not so pronounced in the heart. Measurements of the 3H activity in the plasma after intraperitoneal injection of 3H -ouabain gave 77% higher values in the K^+ -depleted rats than in the controls. This also suggests that the peripheral binding capacity for 3H -ouabain is reduced and that in the K^+ -depleted state the heart may be exposed to a higher concentration of digitalis glycosides.

At variance with the results obtained using HeLa cells⁷, even modest K^+ loss (and a corresponding rise in intracellular Na^+) is associated with a decrease in the number of 3H -ouabain binding sites in rat skeletal muscle (see Figs 1, 2 of ref. 1). This suggests that the synthesis of new $(Na^+ + K^+)ATPase$ units has ceased, and we are now investigating the possibility that this results from a general impairment of enzyme activity or protein synthesis.

Despite our original expectations, we have been unable to detect any rise in the number of 3H -ouabain binding sites in the muscles of K^+ -depleted mice or rats. It seems that skeletal and heart muscle cells differ from cultured cells, perhaps due to central regulatory mechanisms triggered by the drop in plasma K^+ seen a few days after the onset of the K^+ -depleting regime. The decreased capacity for active $Na^+ + K^+$ transport in muscle may prevent any further reduction in plasma K^+ and delay the development of the muscle paralysis which is the final outcome of prolonged K^+ depletion.

AAGE NØRGAARD
KELD KJELDSSEN
TORBEN CLAUSEN

Institute of Physiology,
Aarhus University,
DK-8000 Aarhus C, Denmark

- Nørsgaard, A., Kjeldsen, K. & Clausen, T. *Nature* **293**, 739–741 (1981).
- Chan, P. C. & Sanslone, W. R. *Archs Biochem. Biophys.* **134**, 48–52 (1969).
- Erdmann, E. & Krawietz, W. *Acta biol. med. germ.* **36**, 879–883 (1977).

- Erdmann, E., Bolte, H.-D. & Lüderitz, B. *Archs Biochem. Biophys.* **145**, 121–125 (1971).
- Bluschke, V., Bonn, R. & Greeff, K. *Eur. J. Pharmac.* **37**, 189–195 (1976).
- Clausen, T., Hansen, O. & Larsson, L.-I. *Eur. J. Pharmac.* **72**, 331–335 (1981).
- Boardman, L., Huett, M. & Lamb, J. F. *J. Physiol., Lond.* **241**, 771–794 (1974).

The active radio galaxy 1413+135

BEICHMAN *et al.*¹ and Bregman *et al.*² present interesting IR and high-frequency radio (4.8–14.5 GHz) results on 1413+135. The latter also give an optical spectrum which suggests that it is a galaxy of redshift $z = 0.26$; no emission features were seen. Curiously Beichman *et al.*¹ imply that it was an 'empty field' until discovered as an IR source by Rieke *et al.*³.

The radio source was identified by Hoskins *et al.*⁴ with a 20 mag galaxy on the Palomar Sky Survey, and a finding chart was given. The identification was based on 408-MHz measurements at Molonglo. Condon *et al.*⁵ obtained an accurate radio position which agrees to 1 arc s with the optical position of Hoskins *et al.*⁴. This confirmed the identification but as Bregman *et al.*² point out, the finding chart of Condon *et al.*⁵ is faulty. Their positional offsets are also misleading. Bregman *et al.*² reconfirm the identification of Hoskins *et al.*⁴ but mistakenly refer to this as a 'PKS identification'.

A radio spectrum covering the range 178–2,695 MHz is given by Murdoch and Hoskins⁶. The low-frequency spectrum peaks at 3 Jy at ~300 MHz falling to 1.9 Jy at 178 MHz and 0.67 Jy at 2,695 MHz (in 1971). Bregman *et al.*² report a strong and variable rise at 4.8 GHz and above and Beichman *et al.*¹ show that the intensity peaks at ~100 GHz. In addition to the highly compact component implied by these results, there is evidently a moderately compact component shown by the low-frequency radio data. The overall radio spectrum is quite typical of BL Lac objects and highly variable QSOs. The object would appear to be a galaxy with a BL Lac object in its nucleus.

H. S. MURDOCH

The Astrophysical Department,
School of Physics,
University of Sydney,
New South Wales 2006,
Australia

- Beichman, C. A. *et al.* *Nature* **293**, 711 (1981).
- Bregman, J. N. *et al.* *Nature* **293**, 714 (1981).
- Rieke, G. H., Lebofsky, M. J. & Kinman, T. D. *Astrophys. J. Lett.* **232**, L151 (1979).
- Hoskins, D. G., Murdoch, H. S., Hazard, C. & Jauncey, D. L. *Aust. J. Phys.* **25**, 559 (1972).
- Condon, J. J., Hicks, P. D. & Jauncey, D. L. *Astr. J.* **82**, 692 (1977).
- Murdoch, H. S. & Hoskins, D. G. *Aust. J. Phys.* **26**, 867 (1973).

BOOK REVIEWS

The ontogeny of phylogeny

David L. Hull

PUBLICATION is an unnatural event in the life of a scientific theory. Prior to publishing, a scientist can work in relative privacy, changing his views as he sees fit. After publication, however, the process is made more difficult by the critical stare of the scientific community. Darwin worked on his theory of biological evolution from 1837 until his death 45 years later. Historians have charted his intellectual voyage month by month, sometimes day by day — the *Beagle*, his decision to develop a theory of transmutation, his reading Malthus, the Sketch of 1842, the Essay of 1844, the receipt of Wallace's paper, the hurried publication of the *Origin*, and the trials and tribulations that followed. Ospovat retraces familiar ground in Darwin's story from 1838 to 1844 and then fills in the last gap — the period between 1844 and 1859.

Historians have ignored this period primarily because nothing much of any interest seems to have happened during these years. By 1844 Darwin had already decided that species evolve by chance variation and natural selection. The essentials of his theory were complete. Ospovat argues to the contrary that Darwin made several fundamental changes in his theory some time in the second half of 1856. Prior to that date, Darwin thought that every species is perfectly adapted to its environment, a belief that he carried over from natural theology. He considered that significant amounts of variation occur only in response to intermittent changes in the environment. As a result, evolution for Darwin at this stage in his development was "punctuational". However, in 1856 Darwin formulated his principle of divergence, a principle that he thought was as important as that of natural selection. According to this principle, variation exists within species at all times, allowing them to adapt to variations within their range, including the presence of other organisms. By a sort of "division of labour", species become subdivided into varieties, and these varieties are incipient species. From the beginning Darwin believed that evolution is in some sense progressive, but only after 1856 could he explain why the biosphere tended to become increasingly complex.

In 1855 A.R. Wallace published a paper on the law which regulates the introduction of new species, a paper that both Charles Lyell and Edward Blyth brought to Darwin's attention. Lyell urged Darwin

The Development of Darwin's Theory: Natural History, Natural Theology, and Natural Selection, 1838–1859. By Dov Ospovat. Pp.301. ISBN 0-521-23818-8. (Cambridge University Press: 1981.) £20, \$39.50.

once again to publish lest he be forestalled, and Darwin began his projected multi-volume *Natural Selection*. In 1857 Darwin and Wallace exchanged letters on Wallace's 1855 paper. Then Wallace's paper "On the Tendency of Varieties to Depart Indefinitely from the Original Type" (1858) arrived, not exactly like a "bolt out of the blue". What if Wallace's first paper had prompted Darwin to publish an "abstract" of his views in 1855? What if Wallace's 1858 paper had not caused Darwin to publish the *Origin*? What if Darwin had instead doggedly persisted in his original plan for his Big Book? Answers to such contrary-to-fact questions in history are extremely risky. Even so, Ospovat suggests that the *Origin* would have been a very different book had Darwin published it before formulating his principle of divergence. Instead of making the principle an integral part of his theory as first presented, Darwin would have been forced to introduce it later as an emendation. Similarly, he might have quietly passed on Wallace's paper for publication instead of agreeing to publish a short piece of his own along with it. If he had, I think that very little would have changed except later commentators would have been precluded from going on at such interminable length about Darwin's churlish behaviour. Darwin would still be viewed as the founder of the modern theory of evolution. The Darwin-Wallace papers of 1858 caused hardly a stir; Wallace's paper published alone is unlikely to have engendered more interest. But what of intellectual justice? From this perspective, Wallace still loses out, but this time to Patrick Matthew who had anticipated both Darwin and Wallace as early as 1831.

Although Ospovat makes the requisite passing remarks about the social, political and religious sources of Darwin's scientific views and concludes with a chapter on the development of Darwin's theory as a social process, his book is almost entirely an internalist history of the very best sort. Ospovat not only chronicles Darwin's conceptual development but also presents plausible reconstructions of the factors that led him to adopt the views he did. Not

all of these factors are narrowly "scientific". For example, the tenacity with which Darwin held on to his belief in perfect adaptation stems at least in part from considerations that anyone, both then and now, would count as theological. Darwin was also influenced by prudential considerations of the sort that play a central role in the politics of science. For example, Ospovat argues that Darwin understates the progressive nature of evolution in the *Origin* for fear of offending Lyell and Huxley. But he does not repeat the usual facile claims about the direct translation of sociopolitical interests into the content of science.

Under normal circumstances, I would conclude this review by remarking that the appearance of *The Development of Darwin's Theory* portends a brilliant career for its author. Sadly, however, shortly after the completion of this manuscript, Dov Ospovat died at the age of thirty three. □

David L. Hull is a Distinguished Professor in the Department of Philosophy at the University of Wisconsin-Milwaukee.

Red for green

H. McAllister

The Biological Aspects of Rare Plant Conservation. Edited by Hugh Synge. Pp.558. ISBN 0-471-28004-6. (Wiley: 1981.) £30, \$84.

DESPITE its title, most of this multi-authored book is very usefully devoted to the ways in which botanists in different nations work out which species and communities are in need of conservation in their own countries. As a result, the differences in the scale of the problem between the impoverished but well-known floras of north-western Europe, the rich, less well-known floras of such countries as South Africa, Australia and India, and the very rich, little-known tropical rain forest floras, are clearly brought out.

In north-west Europe the distribution and frequency of occurrence of most species is fairly well documented, so it is

usually obvious which species are rare and in need of conservation. The habitat requirements of such species are mostly known and several chapters deal with how sufficient autecological information has been obtained to devise successful conservation programmes.

The conservation problems of north-west Europe pale into insignificance in comparison with those of such species-rich areas as the Cape of Good Hope with 1,244 threatened plant species. Very few of these have been studied in any detail and the two species discussed exemplify the difficulties. Field work undertaken to promote their conservation has demonstrated the importance of fire in stimulating germination, but, at least in the case of *Staavia dodii*, all attempts to germinate seed outside the natural field environment — including artificial fire — have failed, making garden cultivation impractical as a conservation measure. The concentration of the few remaining populations of so many species in a small area poses serious management problems when so little is known of their autecology.

However, by far the most critical situation exists in the tropical rain forests where a fair proportion of the species are still undescribed. It can be reliably stated that if the present rate of destruction continues, and it is likely to, many species will become extinct before they can be described. There is no doubt that this is the vegetation type in which most extinctions are taking place and that this is a loss not just to biological diversity but to the resources of the world of value to mankind. The four chapters in Section 2 of the book, "Tropical Forests — The Conservation Priority", are an excellent statement of the problem and how it should be tackled. The only sensible approach is to try to conserve areas of great species diversity and hope thereby to protect the maximum number of species. Quite large areas may be required, as a typical tree species may have an average density of 1 per 5 hectares and be dioecious — though sometimes facultatively apomictic. Dransfield's paper describes how controlled cropping can be compatible with conservation.

One chapter, discussing the conservation of lower plants, points to the similarities with the situation in tropical rain forests — a lack of floras and lack of knowledge of species distributions, most species not being amenable to cultivation, meaning that conservation in their native habitats is the only hope for their survival.

The varied topics covered make this work an excellent and much-needed world picture of the present state of plant conservation, although the quality of the individual contributions is very variable. Unfortunately the book promises a bleak future, both environmentally and economically, if there is no improvement. □

H. McAllister is the Assistant Director of the Botanic Gardens, University of Liverpool.

Slimmer tracts for gut physiologists

G.J. Dockray

Physiology of the Gastrointestinal Tract, Vols 1 and 2. Edited by Leonard R. Johnson. Pp.1,492. ISBN 0-89004-440-6. (Raven: 1981.) Two-volume set \$165.

WHEN the American Physiological Society devoted a section of their *Handbook of Physiology* to the gut in the late 1960s, it needed five volumes containing well over 100 chapters to produce a comprehensive account of the subject. After a period of unprecedented expansion in our knowledge of gut function it comes as a pleasant surprise to find that the present work consists of two volumes and less than half the number of chapters needed over a decade ago. Unfortunately, the feeling of relief is tempered by the fact that this brevity has been partly achieved by a selective approach. There is, for instance, relatively little attention given to the liver, salivary glands, comparative and evolutionary aspects, and control of nutrient intake. The omissions are regrettable, not least because they are inimical to the editors' objective of producing a definitive treatise.

This is not to say, however, that there are not good reasons to justify a relatively economical approach in reviewing the present state of gastrointestinal physiology. Recent advances in this field have dealt very largely with the elucidation of events at the cellular and sub-cellular levels. The present volumes both reflect the recent areas of progress and illustrate how it is possible to explain fundamental aspects of organization in a way that readily reconciles fragmentary and even contradictory work from the past.

A series of chapters at the start of the first volume, and at the end of the second, deals with general aspects of digestive physiology and related topics. The major gut functions — secretion, motility and absorption — are covered in separate sections. In each case there are contributions dealing with functional morphology and molecular mechanisms, and these serve to provide a context for chapters dealing with the integrative aspects of each system. The treatment works well, and in particular makes it easy to see how physiological mechanisms of the gut can be related to the functions of other organ systems. Indeed, it is worth noting that a significant part of our understanding of such basic cellular mechanisms as membrane transport and secretion have been derived from studies on the alimentary tract and associated organs. For this reason these two volumes will be of interest and value to many whose main concerns lie outside the gut, but who nevertheless wish to understand recent developments in gut physiology. The gut chauvinist should not, however, feel neglected since there are, for example,

chapters dealing with such singularly digestive subjects as dietary fibre and gut gas.

The contributors are nearly all well-recognized authorities in their fields and most have taken the opportunity to produce a comprehensive assessment of the present state of their subject. This, as much as anything, should serve to ensure the success and value of the work as a whole, since it is likely to be some time before such a strong team will again feel tempted to review their topics at length and in unison. In the present climate of recession and austerity it would be surprising if there were many who felt inclined to buy this work for themselves; they need feel no guilt, however, about persuading their librarians to make the investment. □

G.J. Dockray is Reader in Physiology at the University of Liverpool.

Cladistics clarified

Alan J. Charig

Phylogenetics: The Theory and Practice of Phylogenetic Systematics. By E.O. Wiley. Pp.439. ISBN 0-471-05975-7. (Wiley: 1981.) £27.75, \$49.90.

WILLI Hennig is rightly regarded as the father of phylogenetic systematics; his book, *Phylogenetic Systematics* (University of Illinois Press, 1966) is the fundamental source of this important approach towards the reconstruction of evolutionary history and the classification of living organisms. It is therefore appropriate that Wiley's new book on this subject should use a photograph of Hennig as its frontispiece. During the last decade of Hennig's life (he died in 1976) the school of thought which he had founded — now loosely referred to as "cladistics" — attracted an ever-increasing volume of attention and comment.

Within the last couple of years, however, the Hennigian school has sprouted the new, very different and even more controversial offshoot, "transformed cladistics" ("natural order systematics" in my terminology). The works of Popper on the philosophy of science undoubtedly exerted a major influence upon this schismatic development during its prolonged gestation, but Popper himself seems to have rejected it subsequently. Indeed, were Father Hennig still alive today he too would have rejected it wholeheartedly; his arch-opponents would have been, not Mayr and Simpson, but leading transformed cladists like Nelson and Patterson.

The approach of the transformed cladists is essentially typological; they

maintain that there is a "natural order" in living things, manifesting itself in their aggregate similarity and not requiring to be interpreted as the result of evolution. They believe that time, evolution, genealogical relationships, the fossil record and the possibility of homoplasy have no part to play in systematics. Thus their philosophy is the absolute antithesis of Hennig's or Wiley's, in which — as is implied by the name phylogenetic systematics — phylogeny occupies a central position. Nothing could be more misleading than the recent pronouncement of a transformed cladist that Hennigian cladistics is merely "transformed cladistics with an evolutionary veneer". Contrary to general belief, the battle today is not between cladists and non-cladists but between transformed cladists on the one hand and, on the other, all those — be they followers of Hennig or of Mayr and Simpson — who maintain that the classification of living organisms should be based, wholly or at least in considerable part, upon the phylogeny.

Wiley is unashamedly an evolutionary or Hennigian cladist; he does not even mention transformed cladistics, for the very good reason that he must have written his book just before the new development made its overt appearance. On p.98 he makes brief reference to the ideas of some workers, *circa* 1977, which foreshadowed the emergence of the new school; there is nothing more. Thus the volume reviewed here, unlike so much recent publication on the subject of cladistics, is concerned solely with Hennigian, i.e. evolutionary or phylogenetic cladistics.

Within that unavoidable limitation the book is first-class and might well be described as the evolutionary cladist's vademecum. The central part of the text is organized into five chapters — "Species and Speciation", "Supraspecific Taxa", "Phylogenetic Trees", "Characters and Reconstruction of Phylogenies" and "Phylogenetic Classification". The introduction which precedes it includes a great number of useful definitions clearly explained (although I myself do not agree with all of them) and discusses briefly the relationship of philosophy to systematics and the various approaches to scientific reasoning. Of the later chapters in the book, the first deals critically with the other major systems of taxonomy; the remainder are concerned with biogeography, specimens and curation (even a few pages on field-work and a couple of paragraphs on computer cataloguing), characters and their quantitative analysis, and publication and the Rules of Nomenclature. Bibliography and index run to over 18 pages each, the former providing an invaluable entrée into the literature.

The whole is arranged, set out and written with admirable clarity in excellent English. Compared with Hennig's exposition, Wiley's book is easy to find one's way around in; a remarkable amount of information and commentary on a

comprehensive range of subjects is skilfully compressed into a limited amount of space, yet the author still manages to illustrate and emphasize his points by the extensive use of practical examples. Typographical errors are few. The book, nevertheless, is not without its faults. Thus the important aspect of cladistic analysis which everyone else (as far as I know) refers to as "polarity" is never called by that name and it took me a long time to track it down under the heading "Auxiliary criteria of phylogenetic apomorphy". Certain sections of the book with a mathematical flavour (for example those on Wagner algorithms and quantitative character analysis) were virtually incomprehensible to me, but that surely reflects my own innumeracy rather than any failing on the part of the author.

One of the most refreshing aspects of the book is that Wiley, although obviously a committed evolutionary cladist himself, does not assume the quasi-evangelical attitude so often adopted by his fellow cladists. (Just for the record, I myself am a firm believer in the rigour of a cladistic approach to phylogeny reconstruction but am strongly opposed to a purely cladistic classification.) Further, even more than the many other cladists who show healthy divergences of opinion on various points of doctrine, Wiley is no slavish follower of the majority; he shows a sturdy independence of thought, sometimes going so far as to recant on earlier statements of his own. Thus, unusually, he now accepts not only the postulation of ancestral species but also the possibility of occasionally recognizing them and of testing their hypothesized relationships.

I could make dozens of comments on this book, many of them highly favourable, but here are three of the other sort. First, perhaps because I am a palaeontologist, I deplore Wiley's uncompromising rejection of phyletic speciation into palaeospecies (allochronous speciation) and his consequent acceptance of each entire lineage as one species, irrespective of the amount of anagenetic change. (I note with pleasure, however, that Wiley — unlike Hennig — does not insist that the next dichotomy must inevitably terminate the species.) Secondly, I think it a great pity that he missed this opportunity to sort out the fearful confusion regarding the terms monophyly, polyphyly, holophyly and paraphyly once and for all — it is really so simple. And thirdly, I disagree with him profoundly on the alleged weaknesses of "evolutionary" (i.e. Simpsonian) taxonomy.

Notwithstanding such comments, I still feel that if a systematist were to be allowed only one work of reference on his desert island, he could not do better than choose Wiley's *Phylogenetics*. □

Alan Charig is Chief Curator of Fossil Amphibians, Reptiles and Birds at the British Museum (Natural History).

The Basel network

J.H. Humphrey

The Immune System, Vols 1 and 2. Edited by C.M. Steinberg and I. Lefkovits. Vol.1. *Past and Future*, pp.440, ISBN 3-8055-3407-8; Vol.2. *The Present*, pp.472, ISBN 3-8055-3408-6. (Karger: 1981.) Vol.1. DM 297, SwFr.248, \$148.50; Vol.2 DM 321, SwFr.268, \$160.50.

THESE two volumes comprise 125 essays written as a *Festschrift* for Niels Kai Jerne. They are a tribute not only to his seventieth birthday but also to his directorship of the Basel Institute of Immunology during its first ten years. The essays range from philosophical consideration of neurological processes and of art, through critical examinations of various current and earlier immunological hypotheses, to summaries of new experimental findings, mainly discussed in relation to their bearing on Jerne's own theories. Some are written by old friends and colleagues from his days in Copenhagen, Pasadena, Geneva, Pittsburgh and Freiburg, before he came to Basel, but the great majority are by colleagues who were members, students, visitors or even advisers at the Basel Institute.

The attraction of working at Basel was very great, since Jerne's policy was to have only a small core of permanent members but a much larger number of able, mostly young immunologists from all over the world for anything from six months to five years. He had a flair for choosing persons with talent who would return to work in lively departments elsewhere, and thus the list of contributors to the books reads rather like a selection from an immunological *Who's Who*. Most of them evidently felt challenged to write something original which would amuse or interest Jerne himself as well as any potential readership.

The outcome is as much as anything a view of the immune system as seen from Basel. Here the focus has been on the basic problems of how the immune system recognizes self from not-self, and how there arises the immense variety of different antibodies and receptors on T-lymphocytes which a frog or a mouse or a human being can make. This has involved study of the structure of monoclonal antibodies (initially from myelomas but later from hybridomas also) and of the mechanisms of their assembly and secretion; development of greatly improved culture methods, which made it possible to measure the number of lymphocytes able to respond to any given antigen, or to make immunoglobulin (Ig) in response to polyclonal mitogens; and, when cloning of DNA and DNA sequence studies became possible, comparison of the genes controlling the synthesis of specific antibodies with each other and with corresponding genes in the fetus, which led

in Basel and elsewhere to the recognition of three separate genes controlling the variable region of Ig. Since similar studies on the receptors of T-lymphocytes are frustrated by the difficulty of identifying, let alone purifying T-cell receptors, the variety of T-cells had to be approached by functional assays *in vitro* (mainly the generation of cytotoxic cells but also of helper and suppressor cells) using the improved culture methods developed in Basel to permit enumeration of single cells by limiting dilution assays. Faith and hope are now being pinned on T-cell clones, with recognizable specificities, as a source of receptors which can be isolated and sequenced, with the aim of eventually identifying the genes controlling them and so discovering their relationship to the genes controlling Ig and the Ir region of the major histocompatibility complex.

When, in 1974, Niels Jerne put forward the network hypothesis (which states that for every specific antibody or T-cell receptor which has a distinct combining site (idiotype), the immune system is likely to contain cells able to recognize that idiotype and, in principle, to respond by making more anti-idiotype — which in turn can stimulate cells making anti-anti-idiotype etc., and so establish an infinite network of interactions), some of his colleagues put it to the test. They found that in the special circumstance when mice respond to a bacterial antigen by making antibodies with predominantly a single idiotype, specific anti-idiotype antibodies could either stimulate or suppress antibody formation against the original antigen. The network hypothesis was thus shown to have at least one real mesh; subsequent work on idiotypes has established several more, and a fair amount of the work going on in Basel (and of the more speculative essays in these volumes) is orientated towards the network hypothesis. However since the Institute has a horizontal structure, in the sense that in principle no scientist can tell any other scientist what to do, aspects of immunology which profoundly affect the practical outcome of antigenic stimulation — such as lymphocyte recirculation and the role of macrophages — are also studied there, even though Jerne thought them irrelevant to the basic questions which he hoped to see answered.

The volumes contain essays on all of these subjects, and even some rather good ones on non-specific factors, such as

complement, by scientific advisers who reckoned that Jerne should recognize that they also are important. The more speculative essays, with few or even no references, are included in the first volume and the harder science in the second.

It may reasonably be asked what a collection of this nature has to offer to persons outside the charmed circle of Jerne's associates. Since the contributors felt challenged to write something original, and since free speculation was encouraged and the personalities of the authors show through in their writing, the collection

provides enough stimulation to satisfy any reasonably knowledgeable immunologist. It would take a keen seeker after enlightenment to read through all of the 900 odd pages, but since the editors have grouped the essays under subject headings, selective browsing is quite possible. I would expect to recognize the striking dust jacket (by Jean Tinguely) on train and aeroplane journeys, and even at the bedside. □

J.H. Humphrey is a Professor in the Department of Medicine at the Royal Postgraduate Medical School, University of London.

Down and beyond the composite particle

M.G. Albrow

Jets of Hadrons. By W. Hofmann. Pp.210. ISBN 3-540-10625-1/0-387-10625-1. (Springer-Verlag: 1981.) DM 62, \$28.20.

THE study of strongly interacting particles (protons, neutrons, pions etc.) and the forces between them is no longer frontier physics. We have passed beyond and down to the next layer. It has become established beyond any reasonable doubt that these so-called particles are not elementary at all, but composites of more fundamental constituents. The main effort in high-energy physics is now directed towards studying these constituents and the forces between them. A vast amount of experimental information is all consistent with a theoretical framework describing these constituents as fractionally charged quarks and antiquarks which bind together by exchanging electrically neutral gluons. These entities seem to be at the same level of elementarity as electrons, muons, neutrinos and photons, yet the latter are freely observed unlike quarks and gluons. One can do an experiment (for example scatter a high-energy muon through a known large angle on a proton) which should produce a quark with a known energy and direction. No quark is seen, always a shower of normal composite particles (hadrons) such as pions, kaons and so on. Try harder by increasing, to the limit of present accelerators, the energy of the muon or the angle of scatter, and the showers become more collimated. They are known as jets of hadrons.

Our theoretical outlook has changed so much and the amount of experimental information has grown so rapidly in recent years that Hofmann's review of the field is extremely welcome. To illustrate this rapid development most of the 400 papers in the reference list have been published in the past four years. In such a situation the author must of course take the risk that the book becomes quickly dated. If it does, I suspect that it will be due to the availability of more (and more precise) experimental data rather than to the concepts and the

formalism becoming outdated.

The book "is intended to give an introductory review to the phenomenon of jets in hadronic final states. . .", and the author succeeds admirably in fulfilling this intention. Starting from the simplest situation of e^+e^- annihilation (where the massive virtual photon produced by the annihilation transforms to a $q\bar{q}$ -pair, which in turn materializes as two jets), he leads the reader through jets in parton models and their development as quantum chromodynamics (the theory of quark-gluon forces) to the more complicated processes of hadron-hadron collisions. A chapter on the old "uncorrelated jet model" and one on "soft" hadronic collisions are useful reminders of the importance of phase space effects and four momentum conservation, and also that while theory has a long way to go in describing the most common collisions there is encouraging progress in that direction.

It is a pity that the author was not prepared to address the question of future developments. Now that proton-antiproton collisions are being studied at CERN at energies nearly ten times higher than previously available at the ISR, and that LEP has received approval to study high-energy electron-positron collisions, a discussion of expectations for future experiments would have been a useful addition. Nonetheless, this review is sufficiently detailed and complete that theorists working in the field will find it a valuable guide to the existing experimental data, and experimentalists will appreciate the exposition of the theoretical ideas. It is, however, written at a level such that graduate students in high-energy physics should not find it too heavy-going, and it will familiarize them with what may well remain a major topic in the subject over the next decade. □

M.G. Albrow, currently at CERN, is a Research Physicist at the Rutherford Appleton Laboratory, Chilton, Oxfordshire.

The second volume of *Elsevier's Dictionary of Botany*, edited by Paul Macura, has recently been published by Elsevier Scientific. (For review of Vol. I, *Plant Names*, see *Nature* 284, 382; 1980.) Volume II, *General Terms*, contains some 10,000 entries of English, French and German terms used in botany and associated disciplines, and includes French, German and two Russian indexes, one numerical and the other alphabetical. The book costs Dfl.260, \$110.75.